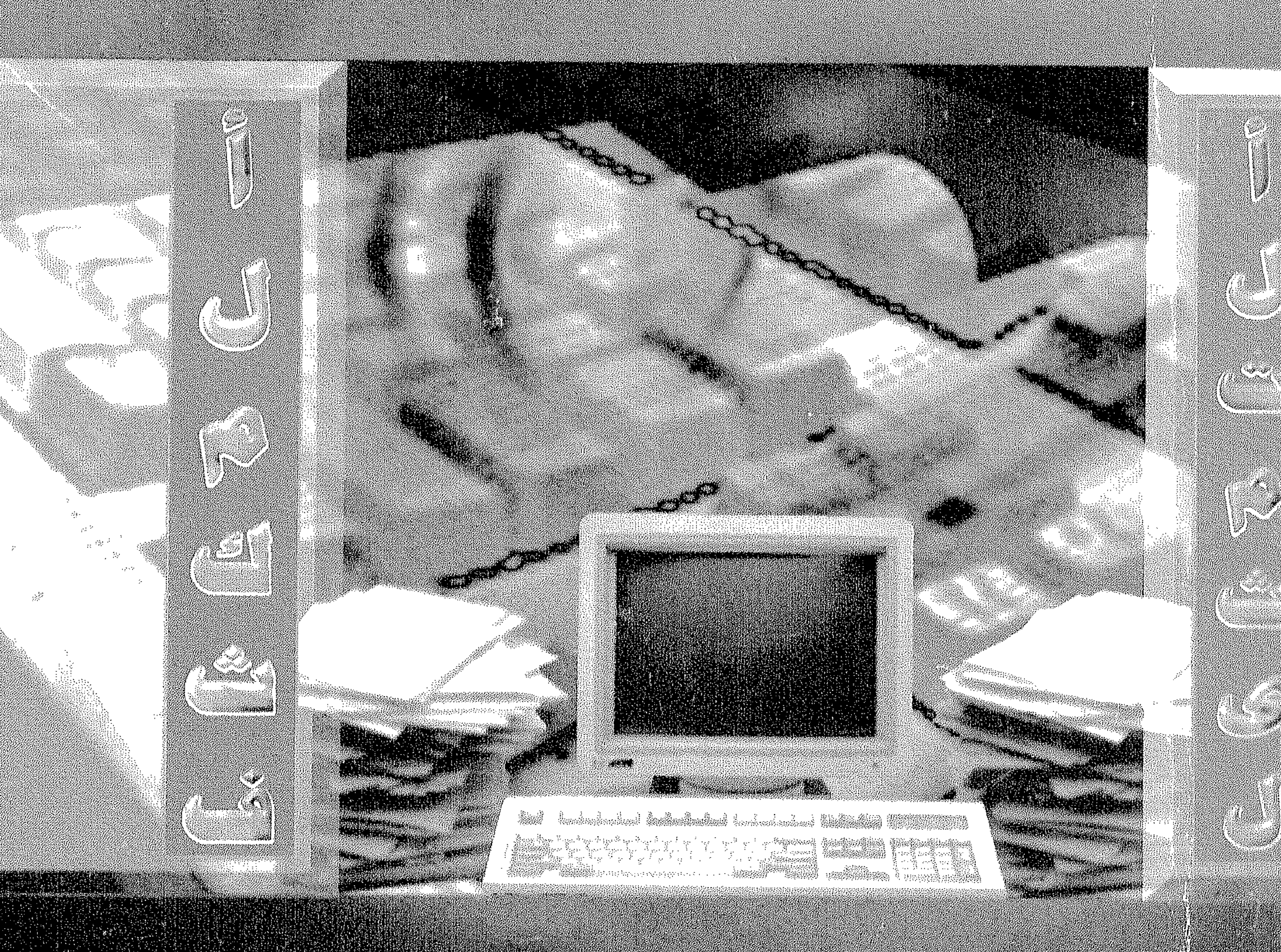
د. أحمد بـدر   د. محمد فتحي عبد الهادى   د. ناريمان إسماعيل متولى

التكشيف والاستخلاص

دراسات فى التحليل الموضوعى



دار قباء للطباعة والنشر والتوزيع



التكشيف والاستخلاص

دراسات فى التحليل الموضوعى



التكشيف والاستخلاص

دراسات فى التحليل الموضوعى

إعداد

د. أحمد بدر
أستاذ المكتبات والمعلومات
جامعة القاهرة

د. محمد فتحى عبد الهادى
أستاذ المكتبات والمعلومات
ووكيل كلية الآداب ـ جامعة القاهرة

د. ناريمان إسماعيل متولى
أستاذ علم المعلومات المشارك
جامعة الإمام محمد بن سعود الإسلامية بالرياض
معارة من جامعة الإسكندرية

الناشر
دار قباء للطباعة والنشر والتوزيع
عبده غريب

الكتـــاب: التكشيف والاستخلاص "دراسات في التحليل الموضوعي"

المؤلـــف: د. أحمد بدر د. محمد فتحي عبد الهادي

د. ناريمان إسماعيل متولي

رقم الإيـداع: ٢٠٠١/١٨١١

الترقيم الدولي: I S B N

977 – 303 – 322 - 8

تـاريخ النشـر : ٢٠٠١

الناشـــــر : دار قباء

للطباعة والنشر والتوزيع

الإدارة :

٥٨ شارع الحجاز – عمارة برج آمون

الدور الأول – شقة ٦

☎ ٦٣٦٢٥٦٢ – فاكس/ ٦٣٧٤٠٣٨

المكتبـة :

١٠ شارع كامل صدقي الفجالة (القاهرة)

☎ ٥٩١٧٥٣٢ / ✉ ١٢٢ (الفجالة)

المطابـع :

مدينة العاشر من رمضان – المنطقة الصناعية (C1)

☎ ٠١٥/٣٦٢٧٢٧

www. alinkya.com/kebaa

e-mail: qabaa@nasecj.com

بسم الله الرحمن الرحيم



محتويات الكتاب

الفصل الثالث

اللغة الطبيعية والمصطلحات المحكومة فى استرجاع المعلومات

الفصل الرابع

التكشيف الآلى: دراسة للتطورات فى المجال خلال أربعين عاماً

الفصل الخامس

تقييم الكشافات والتكشيف

الفصل السادس

المستخلصات والاستخلاص: دراسة مسحية تحليلية

الفصل السابع

التحليل المقارن لمصطلحات ومستخلصات العلوم الطبيعية والاجتماعية والإنسانيات على ضوء معايير الاستخلاص وفى بيئة استرجاع إلكترونية وتقليدية

الفصل الثامن

الاستخلاص الآلى: دراسة للتطورات فى المجال خلال أربعين عاماً مع بعض نتائج دراسات التكشيف والاستخلاص الآلى

الفصل التاسع

تقييم المستخلصات: دراسة فى التطور والمعايير والجودة خلال نصف قرن

الفصل العاشر

المكانز

الأساسيات والاتجاهات الحديثة والإسهامات العربية

الفصل الحادى عشر

الاتجاه المكنزى فى قائمة رؤوس موضوعات مكتبة الكونجرس: ترجمة لمقدمة الطبعة الثالثة والعشرين لعام ٢٠٠٠م

الفصل الثانى عشر

تكشيف واستخلاص الإنتاج الفكرى العربى: دراسة تحليلية

الفصل الثالث عشر

أضواء على التطورات المعاصرة ومستقبل خدمات التكشيف والاستخلاص فى القرن الحادى والعشرين

مقدمة الكتاب

يحتل التحليل الموضوعي بما يشمله من تكشيف واستخلاص وتصنيف وفهرسه موضوعية، مكانا محورياً فى أنشطة المكتبات ومراكز المعلومات على اختلاف مستوياتها وأنواعها، ذلك لأن التحليل الموضوعي يهتم ببناء التمثيلات الموضوعية representations اللازمة لتحليل المواد المنشورة فى قواعد البيانات، سواء كانت هذه المواد مطبوعة أو فى شكل مقروء آلياً، أو فى شكل بطاقى أو على الانترنت (الوب) أو غير ذلك.

كما يهتم التحليل الموضوعى بالتمثيل المكثف الموجز والدقيق لمحتويات الوثائق من أجل إعداد المستخلصات، وهى التى تحتوى على بعض المعلومات الفعلية التى تحتويها الوثيقة، وتلعب المصطلحات الكشفية دوراً هاماً فى بناء هذه المستخلصات.

وإذا كانت عمليات التحليل الموضوعى تتكامل فيما بينها، فيركز هذا الكتاب على دراسات أعدها – أو حدثها – المؤلفون الثلاثة فى مجالات التكشيف والاستخلاص، ونشر معظمها منذ بداية الستينيات، وحتى بداية القرن الواحد والعشرين.

هذا ويهدف الكتاب إلى التركيز على الجوانب التالية:

(أ) يشتمل الكتاب على المعلومات القاعدية الأساسية فى التكشيف والاستخلاص، فضلاً عن التطورات والبحوث الجارية في المجال، أى أن الكتاب يتوجه إلى الطالب فى مراحله الجامعية الأولى فضلاً عن الباحثين فى مستوى الدراسات العليا.

(ب) حرص المؤلفون على تقديم الدراسات والبحوث حتى عام ٢٠٠٠م حتى يعكس الكتاب الإنتاج الفكرى الحديث والمعاصر فى المجال، وقد وصلت المراجع الإنجليزية إلى أكثر من (١٧٥) مرجعاً منها نسبة حوالى ٧٥% صدرت فى الثمانينيات والتسعينيات.

(ج) هناك ترابط شديد بين مختلف عمليات التحليل الموضوعى، فقائمة المصطلحات التى تمثل مجالاً معيناً، يمكن اعتبارها نوعاً من المستخلص

المصغر، والمكنز عادة يربط بين التكشيف والتصنيف، وحتى رؤوس الموضوعات فى ثوبها المتطور تتجه نحـــو التركيبات المكنزية.

والكتاب الذى بين أيدينا يعكس هذا الترابط بوضوح **فالفصل الأول** يعكس البدايات التاريخية للتكشيف والاستخلاص ودورهما فى نجاح إجراءات البحث واسترجاع المعلومات، فضلاً عن التعرف على مختلف التعاريف والعمليات المتصلة بالتحليل الموضوعى، أما **الفصول الأربعة** التالية فهى تتناول التكشيف والكشافات من حيث مستوياتها وأنواعها وخطوات عمليات التكشيف وخصائصها، ومن حيث المقارنة بين اللغة الطبيعية والمصطلحات المحكومة ومن حيث التكشيف الآلى وأهميته المستقبلية ثم تقييم الكشافات والتكشيف ..

أما **الفصول الأربعة** التى تليها فتتناول المستخلصات والاستخلاص وتطبيقات معاييرها والتحليل المقارن لمستخلصات العلوم الطبيعية والاجتماعية ثم الاستخلاص الآلى وكيفية تقييم المستخلصات والتعرف على مدى جودتها.

أما **الفصول الأربعة** المتبقية فهى تركز على المكانز والاتجاهات المكنزية لرؤوس الموضوعات، ثم تحليلاً للإنتاج الفكرى العربى فى مجال التكشيف والاستخلاص وأخيراً إلقاء الأضواء على التطورات فى المجال والتوقعات الخاصة بمستقبل الخدمات التقليدية فى التكشيف والاستخلاص فى القرن الحادى والعشرين.

(د) يمثل المؤلفون الثلاثة أجيالاً ثلاثة فى دراسات المعلومات والمكتبات الأول من جيل الرواد (الجذور) الذى بدأ تدريس التوثيق العلمى بجامعة القاهرة فى منتصف الستينيات، وكان التكشيف والاستخلاص واحدا من اهتماماته التدريسية ومقالاته المنشورة، وقد كتب الفصول الستة التالية (الأول، الثانى، الثالث، الخامس، السابع والحادى عشر) أما المؤلف الثانى فهو من الجيل الثانى (الجذوع) ولكنه يعتبر رائد التأليف والنشر فى الوطن العربى فى مجال التكشيف وخصوصاً المكانز وقد كتب الفصلين العاشر والثانى عشر أما المؤلفة الثالثة فهى من الجيل الثالث وقد كتبت ونشرت الفصول الأربعة التالية (الرابع/السادس/الثامن/ التاسع) وشاركت مناصفة مع المؤلف الأول فى إعداد الفصل الثالث عشر الأخير.

والله نسأل أن يكون هذا الكتاب إضافة للكتب العربية القليلة فى الإنتاج الفكرى المعلوماتى والتى تتناول موضوع التكشيف والاستخلاص، وأن يكون الكتاب معيناً للطلاب وأعضاء هيئة التدريس على المستوى الجامعى الأول وعلى مستوى الدراسات العليا والله من وراء القصد .

المؤلفون

١٢/١٢/٢٠٠٠م

الموافق ١٦ رمضان ١٤٢١هـ



الفصل الأول

التكشيف والاستخلاص: دراسة
في التاريخ والتعاريف والأهمية ومشكلة استرجاع المعلومات

يتناول هذا الفصل تعريفاً بالبدايات التاريخية لعمليات التكشيف والاستخلاص وتطورها من الحضارات القديمة وحتى القرن العشرين، ثم بيان علاقة التكشيف والاستخلاص بالتحليل الموضوعي والتمييز بين مصطلحات التكشيف الموضوعي والفهرسة الموضوعية والتصنيف ثم التعريف بالأهمية المتزايدة للإفادة من الحاسبات الآلية في خدمات التكشيف والاستخلاص بطبيعة المعلومات وإجراءات البحث ودور التكشيف والاستخلاص في نجاح هذه الإجراءات.

أولاً: بدايات تاريخية

١- مقدمة:

تاريخ التكشيف والاستخلاص قديم قدم الحضارة الإنسانية، فقد حفظت أقراص الطين في حضارة بابل وأشور، وكتب عليها من الخارج المحتويات التي بداخلها، كما حفظت أوراق البردي في حضارة مصر القديمة — في أوانٍ كتب عليها كلمات تدل على محتواها (Borko, H. 1975) .

ولعل أهم المحاولات الرسمية الببليوجرافية تعود إلى كاليماكوس في مكتبة الإسكندرية وكان ذلك قبل ميلاد المسيح بحوالي قرنين من الزمان .. وكان كاليماكوس يعد قوائم مرشدة عن المعلومات الخاصة بآلاف لفائف البردي الموجودة بالمكتبة وفي الواقع فقد عاشت العديد من المستخلصات حتى وقتنا الحاضر بينما اندثرت المصادر الأصلية .. أي أن المستخلصات تعتبر المصدر الوحيد أمام الباحثين في كثير من الأحيان.

وفي خلال هذه الفترة قام اليونان القدماء باستخلاص عقد الروايات Plots of plays ثم إدخالها قبل النص وفي بدايته مع قائمة الممثلين .. وكانت هذه المستخلصات تسمى الفروض Hypotheses باليونانية .. ولقد وجد الباحثون

المحدثون أن هذه الأدلة المرشدة للمحتويات، تعتبر كأدوات أساسية فـــى إعـــادة فهم الأحداث اليومية الاجتماعية والاقتصادية القديمة.

هذا ولم تكن المستخلصات تستخدم فى الأعمال البحثية وحدها، ذلك لأنه فى القرن الثالث قبل الميلاد كانت هناك تسجيلات فى مجلات الأعمال وغيرها .. مصحوبة بمستخلصات دالة على محتوياتها (Cleveland, D., 1990).

هذا وقد تضمنت الأعمال غير الروائية القديمة رؤوس الفصول أو ملخصات والتى تعبر عن المحتوى الذى يأتى بعد ذلك، ولعل الإنجيل فى القرون الأولى كان يعبر عن ذلك تعبيراً ملائماً .. ولكن كشافات الإنجيل وكشافات النصوص Concordances جاءت بعد ذلك بوقت طويل.

وهناك بدايات أيضاً للتكشيف فى الحضارة العربية الإسلامية وذلك بالنسبة للقرآن الكريم أو الأحاديث النبوية الشريفة، حيث توضع بعض الكلمات المفتاحية على الجانب الأيمن للنصوص للدلالة على محتوياتها.

هذا وفى الوقت الذى اخترعت فيه الطباعة المتحركة على يد جوتنبرج حوالى عام ١٤٥٠م كانت هناك بعض أشكال الحواشى Annotations وهى:

(أ) حواشى المخطوطات. (ب) فهارس المكتبات (ج) الببليوجرافيات

كما احتوت الكتب فى بداية عصر الطباعة على صفحات بيضاء فى بداية ونهاية الكتاب لكى يسجل عليها القراء الموضوعات ذات الأهمية التى يحتويها الكتاب حتى يمكن الرجوع إليها بعد ذلك.

ويعنى ذلك أن الكشافات قديمة قدم الكتابة المنشورة، وقد وجدت فى الواقع فى جميع اللغات، وإن اقتصرت الكشافات الأولى على أسماء الأشخاص أو على كلمات النص أى أنها لم تكن كشافات موضوعية، ولكن التكشيف فى مفهومه وشكله المعاصر لم يبرز حتى استقر مفهوم الترتيب الهجائى Alphabetizing ..وطبقاً لما يذهب إليه الباحث وتى (Witty, 1973) .. فإنه على الرغم من استخدام الترتيب الهجائي فى العصور اليونانية الرومانية ووضع الملخصات عند بداية بعض الأعمال غير الروائية. فليس لدينا أى شىء شبيه بالكشاف الهجائى

قبل العصور الوسيطة.. كما أن التكشيف الهجائى لم يظهر فى أرشيف الفاتيكان قبل القرن الرابع عشر الميلادى .

وهناك بعض الاستثناءات بالطبع ففى حوالى عام ٩٠٠ بعد الميلاد كانت هناك موسوعة واحدة مرتبة هجائياً ولكن معظم هذه الأعمال كانت مصنفة .. وحتى بعد ظهور الكشافات الحقيقية، فقد كانت طويلة، كما كانت الكشافات اللاتينية القديمة مرتبة ببساطة حسب تجميع المداخل تحت الحروف، ولكن دون ترتيب معين داخل هذه الحروف، ولم تنتنه تلك الممارسات حتى القرن السابع عشر (Cleveland, D., 1990, 4) هذا وقد كانت كشافات الكتب تعد بواسطة المؤلفين حتى القرن الثامن عشر، كما شهد هذا القرن كذلك دخول المكشف المهنى إلى الميدان، ولكن المحاولات الهادفة لتجميع الكشافات التى تغطى جميع مجالات المعرفة، كانت فى القرن التاسع عشر.

أى أن القرن التاسع عشر قد شهد تصنيف Taxonomy عمل المكتبات والتوثيق، وأصبحت الكشافات الموضوعية منتشرة وأكثر منهجية، كما زادت آلية الطباعة وتطورها من تحقيق الحاجة إلى الكشافات، وظهرت كشافات الدوريات، وبدلاً من كشاف واحد للدورية فى العام الواحد، أدخل بول Pool كشافه الذى يضم عدة دوريات، والذى يغطى فى نفس الوقت عدة سنوات.

أما ولسن Wilson فقد بدأ مرشد القارىء للإنتاج الفكرى فى الدوريات عام ١٩٠١م — Reader's Guide to periodical literature, 1901

وقد اشتهر هذا الكشاف بمداخله الموضوعية وإحالاته المتميزة، وجاء تكشيف مقالة الدورية هنا تحت اسم المؤلف وتحت الموضوع المحدد، وكانت هناك إحالات عديدة لربط كل موضوع بالموضوعات المتعلقة به.

ولم يكن هذا الكشاف هو أول عمل ظهر فى العصر الحديث، بل سبقه كشاف الإنتاج الفكرى للدوريات عام ١٨٨٢م، كما أن أول مجلة استخلاص نشرت فى يناير ١٦٦٥م تحت اسم Journal des Scavans أى مجلة العلماء، وقد خصصت نصف كل صفحة لمادة واحدة، أى لكتاب جديد مثلاً مع تفصيلات عن مؤلفه وعنوانه ومكان نشره.

وقد شهدت السنوات التى تلت عام ١٧٠٠م بداية ظهور دوريات الاستخلاص فى ألمانيا، ثم فى انجلترا فى القرن الثامن عشر ثم فى فرنسا، ومعظم هذه الدوريات كانت شبيهة بمجلة العلماء التى سبقت فى عام ١٦٦٥م.

وعلى كل حال فأمناء المكتبات لم يهتموا بدخول مجال كشافات الدوريات إلا عندما أحسوا بالحاجة إلى مثل هذه الكشافات وكان ذلك واضحاً فى اجتماع الجمعية الأمريكية للمكتبات (ALA) عام ١٨٧٦م. وقد أدت المناقشات فى هذه الاجتماعات إلى جهود تعاونية بين المكتبات والتى أدت بدورها إلى ظهور كشاف الإنتاج الفكرى للدوريات عام ١٨٨٢م Index to periodical literature وشهد القرن التاسع عشر (فى نهايته) اهتماماً بالتكشيف مع ظهور حاجة الباحثين إلى تحسين الضبط الببليوجرافى للإنتاج الفكرى، وقد تأسس المعهد الدولى للببليوجرافيا على يد بول أوتليت وهنرى لافونتين عام ١٨٩٢م وكانت من بين اهتماماتهم مشكلة الإتاحة بالنسبة للمعرفة المسجلة، ولقد تطورت هذه الإتاحة الموضوعية عن طريق خطط التصنيف، ولم يكن الباحثون راضين عن هذه الخطط العامة بل كانوا فى أمس الحاجة إلى المعلومات المحددة.. وقد اقترحت خطط بديلة كتكشيف كلمات العنوان وهذه أدت بدورها إلى الكلمات المفتاحية المعاصرة أو التكشيف الحر للنص Free-text indexing.

وإذا كان عام ١٦٦٥م قد شهد مولد أول دورية استخلاص، فقد كان القرن الثامن عشر (١٧٠٠-) مشيراً إلى ظهور دوريات استخلاص أخرى ففى عام (١٧٠٣م) ظهرت دورية المستخلصات الألمانية Monat Sextracte والتى بدأت فى مدينة ليبزج Leipzig، كما بدأت فى القرن الثامن عشر فى انجلترا وفرنسا دوريات مشابهة .. وكانت ساحة التبادل الفكرى بين العلماء المنتشرين فى أوروبا.

ولقد بدأت الدوريات الاستخلاصية فى التخصص منذ القرن التاسع عشر وذلك استجابة لزيادة عدد الدوريات الأولية Primary journals كما تحسنت أساليب تكشيف الكتب، وزادت خبرات التكشيف المكتسبة عبر السنين.

هذا وقد شهد القرن العشرون أكثر الكشافات الموضوعية انتشاراً والتى احتوت على قائمة المصطلحات دون أقسام فرعية أو أنها احتوت على رؤوس فرعية sub

divisions فى بناء هرمى بسيط لمستويين أو ثلاثة مع وجود المصطلحات أو الجمل على كل مستوى مرتبة فى ترتيب هجائى.

ولقد حدث تغير ملموس فى اتجاهات الناشرين والأمناء والمستفيدين من المعلومات وذلك بالنسبة للاهتمام الزائد بعمليات التكشيف والاستخلاص، ولعل ذلك يعود بالدرجة الأولى لفيضان أو إنفجار المعلومات كما يطلق على ذلك أحياناً وذلك بعد الحرب العالمية الثانية.

ولقد تطورت أساليب التكشيف بسرعة كبيرة واستخدمت الآلات لإنتاج وفرز وبحث بطاقات التكشيف بل وفى إنشاء الكشافات جميعها، كما وضعت نظم الطباعة وتوزيع الكشافات بطريقة أكثر سرعة وكفاءة.

هذا وقد شهدت الخمسينات جهوداً عديدة لتقييم لغات التكشيف باستخدام الطرق الكمية، ولعل مشروعات كرانفيلد (الأولى والثانية) تعتبر علامات بارزة ورائدة فى تطوير أساليب التقييم الكمى للغات التكشيف.

ولقد تجاوز التكشيف فى القرن العشرين أفكار العصور السابقة فمن الكشافات الخاصة بالأعمال الفردية إلى كشافات المجلدات المتعددة إلى الكشافات التعاونية على المستوى الدولى. وهناك فى الوقت الحاضر كشافات على المستوى الدولى فى مجالات الشئون الدولية والتعليم وغيرها من الموضوعات، وإذا كان الإنتاج الفكرى العالمى مازال بعيداً عن التكشيف والاستخلاص الكامل، فقد سار العالم أشواطاً طويلة إلى الأمام.

٢- القيمة المضافة للتكشيف والاستخلاص فى المجتمع المعلوماتى الجديد:

المعلومات معنا منذ بداية التاريخ الإنسانى شفوية أو مكتوبة، مخطوطة أو مطبوعة أو مسجلة على الوسائل الإلكترونية الحديثة. وتقع المعلومات فى قلب التطور الزراعى والصناعى والإدارى والصحى. بل فى قلب النشاط الإنسانى كله .. وإذا كانت التكنولوجيا الأولى هى اختراع الإنسان للغة فقد كانت خطوته العملاقة التالية هى اكتشاف طرق جديدة لتسجيل أفكاره حتى يمكن للمعلومات أن تنتقل عبر الزمن وعبر المكان لتصل إلى كل إنسان على هذه الأرض.

وإذا كانت المكتبات التقليدية قد قامت بحفظ المعلومات عبر العصور، فقد شهد العالم منذ الحرب العالمية الثانية تطوراً ضخماً فى استثمار المعلومات خصوصاً مع التطورات التكنولوجية المتلاحقة .. ومع ذلك استمرت مشكلات تحليل المعلومات قائمة بالنسبة للتكشيف والتصنيف وإجراءات البحث وغيرها.

لقد أصبح للمعلومات فى وقتنا الحاضر دوراً قومياً، وأصبحت صناعة المعرفة والمعلومات من بين الصناعات ذات النسبة العالية فى إجمالى الناتج المحلى (G D P) ودخلت العديد من الدول المتقدمة ما يسمى بمجتمع المعلومات، ذلك لأنه إذا كان رأس المال هو المصدر أو المورد الاستراتيجى فى المجتمع الصناعى، فالمعلومات هى المورد الاستراتيجى للمجتمع المعلوماتى وكما يذهب نياسبيت .Naisbitt,, J) (1984, 11 فإن هناك خمسة أشياء تعكس التحول من المجتمع الصناعى إلى المجتمع المعلوماتى وهى :

(أ) المجتمع المعلوماتى هو حقيقة اقتصادية وليس تجريد ذهنى أو فكرى.

(ب) الاختراعات فى مجالات الاتصالات والحاسبات سوف تعجل من سرعة التغيير لأنها ستقلل الفترة الزمنية بين مرسل المعلومات ومستقبلها فى عملية الاتصال.

(ج) ستطبق تكنولوجيا المعلومات الجديدة على الأنشطة الصناعية القديمة فى البداية ثم يتم توليد أنشطة وعمليات ومنتجات جديدة.

(د) هذا المجتمع فى حاجة أشد للمهارات الأساسية فى محو الأمية المعلوماتية وستكون مخرجات تظامنا التعليمى منتجات متدنية أمام متطلبات العصر.

(هـــ) لا تعتبر تكنولوجيا العصر المعلوماتى الجديد مطلقة، فهى ستنجح أو تفشل طبقاً لمبدأ تكنولوجيا عالية/ لمسة عالية. High Tech/High Touch

وستصبح المعلومات شأنها فى ذلك شأن جميع الموارد الاقتصادية مكلفة فى إنتاجها وبالتالى سيقوم الناس بدفع ثمنها .. ومع تدفق المعلومات الكمى وتعقدها الموضوعى فإن مهنة المعلومات ستتحمل عبء الاسترجاع الجيد وبالتالى فإن نظم التكشيف والاسترجاع الكافى والمرضى للمستفيدين سيصبح أمراً حاسماً.

وما ينبغى أن نتذكره كذلك دخول خدمات ونظم المعلومات فى الثروة القومية ضمن قطاع المعلومات بالدولة، أى أن هذه الخدمات بما فيها من نظم التكشيف والتصنيف والاستخلاص تضيف قيمة Add-Value للمعلومات والبيانات الخام وتجعلها معرفة أى أصبح لخدمات ونظم المعلومات دوراً اقتصادياً هاماً. (Taylor, R., 1984).

ثانياً: فى المصطلحات والعلاقات

١- التحليل الموضوعى وعلاقته بالتكشيف والفهرسة والتصنيف

يذهب لانكستر إلى أن التحليل الموضوعى يعنى التحديد والتعبير عن المحتوى الموضوعى فى النصوص الوثائقية أو فى قواعد البيانات أو فى اللغات المحكومة أو الطبيعية أو فى الطلبات المعلوماتية واستراتيجية البحث .. والتحليل الموضوعى من وجهة نظر المستفيد، محكوم بشدة بالإتاحة الموضوعية، وبالتالى فإن الأساليب التى يمكن بواسطتها استرجاع المعلومات عن موضوع معين من مصادر مختلفة تعتبر أيضاً ذات ارتباط وعلاقة بالإتاحة (Lancaster, F. 1991).

أما الباحث شان وزملاؤه (Chan, L. M. et al 1985) فيذهبون فى مقدمة كتابهم وهو بعنوان "التحليل الموضوعى" إلى أن التحليل الموضوعى يعنى بناء المصطلحات والتكشيف الموضوعى، والمقصود ببناء المصطلحات، بناء الأدوات كالتصانيف وقوائم رؤوس الموضوعات والمكانز وهذه الأدوات تصمم لتسهيل تنظيم واسترجاع المعلومات عن طريق:

أ – وضع مصطلحات يمكن استخدامها فى تصنيف أو تكشيف الوثائق.

ب – عرض العلاقات الدلالية Semantic كالهرمية Hierarchy والترادفية Synonymy بين المصلحطات.

أما بالنسبة لنظرية التكشيف .. فقد استخدام لانكستر هذا المصطلح بينما استخدم باحثون آخرون مصطلحاً أكثر شمولاً وهو نظرية التحليل الموضوعى (Chan, L. et al, 1985) ، إلا أننا مازلنا فى نظر الباحث بوركو (Borko, H., 1977) فى التسعينيات نحاول الوصول إلى نظرية كما كان الحال فى السبعينيات.

وقد ذهب الباحث سفينيونس (Svenonius, 1981, 91) إلى أنه فى غياب نظرية لاسترجاع المعلومات، فليس هناك نظرية واضحة المعالم للتكشيف، وحتى مع وضع المعايير الخاصة بالتكشيف، فإن هذه المعايير تهتم بالممارسة وليس بالنظرية، وإن كان سيفنيونس وآخرون قد عملوا منذ بداية الثمانينيات على بناء نظرية للتكشيف كأساس لتحديد ماذا تقوله الوثيقة Aboutness من معلومات .. خصوصاً والبحث على الخط المباشر يقدم لنا بيئة غنية للمضى فى هذا الطريق.

وعلى كل حال ففى غياب النظرية هناك اختلاف أيضاً بالنسبة لتعريف التكشيف فقد كان وما يزال هناك تمييز بين الفهرسة الموضوعية والتكشيف الموضوعى .. وقد أصبح هذا التمييز فى الممارسة غير واضح، وإن كانت الفهرسة الموضوعية قد ارتبطت بالوصف الموضوعى للوثائق كما هو الحال باستخدام قوائم رؤوس موضوعات مكتبة الكونجرس بينما يرتبط التكشيف الموضوعى بأدوات التكشيف التى وضعتها هيئات تجارية خارج المكتبات .. والتكشيف الموضوعى ارتبط بتكشيف الدوريات أكثر من أى وعاء آخر.

كما أن التكشيف الموضوعى يعنى تطبيق المصطلحات لإظهار محتوى الوثائق أو ما تتحدث عنه aboutness .. ويلاحظ أن التكشيف الموضوعى يحدد مجاله تقليدياً فى بعض أشكال الوثائق كالفقرات داخل الكتب (الكشاف فى نهاية الكتاب) أو مقالات الدوريات. فضلاً عن أن التكشيف الموضوعى يعبر عن بعض أشكال الخيط أو التسلسل String كالواصفات ومصطلحات التكشيف بالمقابلة برؤوس الموضوعات أو أرقام التصنيف، ومع ذلك فيمكن القول بصفة عامة وطبقاً للمفاهيم الأكثر حداثة أن التكشيف الموضوعى يهتم ببيان الفكرة العامة Theme أو الموضوع الخاص بأى وثيقة، وذلك بأى خيط له دلالة ويحمل الصفات الألفبائية أو الرقيمة.

والقيمة التى يمكن أن نجنيها بوضع المعنى العريض للتحليل الموضوعى هى:

أ – يمكننا ذلك من مقارنة عدة مداخل فى التحليل الموضوعى.

ب– يمكننا من التعميم بالنسبة لهذه المداخل على مستوى وصفى عال نسبياً حتى تظهر القواعد والأهداف فى يسر.

ج – يمكننا من وضع نظرة موحدة للمداخل التقليدية والعلمية المعلوماتية للتحليل الموضوعى.

أما لانكستر فى كتابه عن التكشيف والاستخلاص (Lancaster, F., 1991, 14-16) فقد ميز بين ثلاثة مصطلحات وهى:

التكشيف الموضوعى Subject Indexing والفهرسة الموضوعية Subject Cataloging والتصنيف .. فالفهرسة الموضوعية تدل عادة على تخصيص موضوعات تمثل المحتويات الشاملة Overell contents للأعمال الببليوجرافية الكاملة (كتب، تقارير، دوريات ... إلخ) وذلك داخل فهرس المكتبة. أما التكشيف الموضوعى فهو مصطلح يستخدم بطريقة أكثر مرونة، فهو قد يدل على تمثيل الجانب الموضوعى لأجزاء من الأعمال الببليوجرافية الكاملة. وذلك كما هو الحال عند إعداد كشاف فى نهاية الكتاب .. وعلى ذلك فقد تفهرس المكتبة كتاباً تحت رأس موضوع "القطط" للدلالة على الجانب الموضوعى الشامل للكتاب، أما المحتويات التفصيلية لهذا الكتاب فيتم تكشيفها فقط فى نهاية الكتاب عند الكشاف الموضوعى. ولكن يجب أن نلاحظ أن هذا التمييز بين مصطلح "الفهرسة الموضوعية" والتكشيف الموضوعى حيث يدل المصطلح الأول على الأعمال الببليوجرافية الكاملة ويدل المصطلح الثانى على أجزاء من تلك الأعمال .. هذا التمييز اصطناعى ولعله تمييز مضلل .. ذلك لأن العملية التى يتم بواسطتها تمثيل الجانب الموضوعى للأعمال الببليوجرافية فى قواعد البيانات المنشورة ــ المطبوعة أو فى شكل مقروء آلياً ــ يتم الدلالة عليها عادة بمصطلح التكشيف الموضوعى سواء كان ذلك لمعالجة الأعمال بشكل شامل أو لمعالجة أجزاء منها فقط. فنحن مثلاً حين نفحص الكشاف الموضوعى للمستخلصات الكيميائية (Chemical Abstacts) فسنجد هذا الكشاف يدلنا على كتب كاملة أو تقارير فنية كاملة فضلاً عن أنه يدلنا على أجزاء من الأعمال الببليوجرافية (فصول من كتب أو أوراق فى أعمال المؤتمرات أو مقالات فى دوريات).

وعلى الجانب الآخر فقد تختار المكتبات تمثيل أجزاء من الكتب (بعض الفصول أو أوراق البحوث مثلاً) داخل الفهرس، وهذه يشار إليها عادة بأنها فهرسة تحليلية (أى كجزء من الفهرسة الوصفية). وعندما يطبق نفس المبدأ على الجانب الموضوعى فهذا النشاط سيكون فهرسة موضوعية تحليلية.

أما بالنسبة للتصنيف فيميل الأمناء إلى استخدام الكلمة للدلالة على تخصيص رموز أو أرقام تصنيف (من الخطط العامة كخطة تصنيف ديوى العشرية أو خطة تصنيف مكتبة الكونجرس أو غيرها) للأعمال الببليوجرافية وذلك بغرض ترتيب هذه المواد على رفوف المكتبة ولكن الفهرس الموضوعى للمكتبة قد يكون هجائياً (يطلق عليه الفهرس الموضوعى الهجائى) أو أن يرتب طبقاً للتسلسل المنطقى للرموز أو الأرقام. بخطة التصنيف (ويطلق عليه الفهرس المصنف).. والحقيقة هى أن التصنيف – فى معناه الواسع – يتداخل فى جميع الأنشطة المرتبطة باختزان واسترجاع المعلومات، ذلك لأن التصنيف هنا يدلنا على تكوين أقسام من المواد على أساس محتواها الموضوعى. فالمكانز وقوائم رؤوس الموضوعات وخطط التصنيف الببليوجرافية هى فى جوهرها قوائم يتم بواسطتها تحديد هذه الأقسام Classes وترتيبها .. كما أن عملية البحث فى الإنتاج الفكرى تتضمن تحديد أى هذه الأقسام Classes يمكن استشارتها فى الكشاف المطبوع أو الفهرس البطاقى أو فى قاعدة البيانات المقروءة آلياً .. أى أن التصنيف يشكل الأساس الفكرى فى عملية التحليل الموضوعى. والمصطلحات والأنشطة الأخرى داخل التحليل الموضوعى تتكامل فيها بينها لتقديم المحتوى الموضوعى للمواد الببليوجرافية والتعبير عنها بالكلمة أو الرمز أو الرقم.

أما الباحث كليفليند (Clevelend, D., 1990, 16) فيعقد مقارنة عملية وفكرية بين التكشيف والتصنيف، حيث يعتبر أن كلاً منهما عملية فكرية واحدة تتمثل فى تحديد الموضوع أو الموضوعات التى تحتويها وثيقة معينة. والتعبير عن ذلك برقم التصنيف أو بكلمة وهى مصطلح التكشيف .. وفي حقيقة الأمر فإن التكشيف هو عملية تصنيف، فنحن حين نضع مصطلحاً كشفياً لوثيقة معينة تحت "القطط" فنحن بذلك نضع هذه الوثيقة فى فئة أشياء Class of Objects لها صفات القطط.

وإذا كان التصنيف هو الترتيب المنهجى للأشياء وتجميعها فى فئات طبقاً لمعايير معينة، فإن الكشاف يعتبر قائمة للمعلومات الببليوجرافية أو تكشيف استشهاد (Citation) لجسد معين من الإنتاج الفكرى، ويرتب الكشاف عادة بطريقة هجائية اعتماداً على المؤلف أو الموضوع أو الكلمة المفتاحية..

هذا ويعتبر التصنيف نشاطاً طبيعياً للعقل الإنسانى، ذلك لأن الطيف الكامل لعملياتنا الفكرية تعتمد على التصنيف لتنظيم انطباعاتنا التى تستقبلها عقولنا عن طريق الحواس، ثم نحن نقوم بعد الاستقبال مباشرة بمحاولة فهم قطعة جديدة من المعلومات عن طريق العثور على شىء مماثل فى إطار معرفتنا وتجربتنا .. ونظراً لأن التصنيف يوضح لنا العلاقات بين الوحدات الأساسية وبين مجموعات الوحدات Classes، فنحن نستطيع فى هذه الحالة الربط بالمعلومات الجديدة التى نقوم بتجهيزها ومعالجتها.

ومن الطبيعى بناء على ذلك أننا نلجأ للتصنيف عندما نحتاج إلى وضع آلية للضبط الببليوجرافى، ذلك لأن الهدف منه هو معاونة المستفيدين فى تحديد أى جزء صغير من قاعدة المعرفة المختزنة الهائلة والتى تستجيب لاحتياجاتهم فمن بين المليون وثيقة يمكننا بسرعة ودقة تحديد العشرين أو الثلاثين وثيقة التى نحتاجها .. وأفضل وسيلة للوصول إلى ذلك هو تنظيم الوثائق بحيث تظهر لنا العلاقات المعلوماتية بينها ثم توجيه المستفيد إلى المجموعات التى يحتاجها .. وهذا هو هدف كل من التصنيف الببليوجرافى والتكشيف الموضوعى.. أى أن الاثنين متشابهان من ناحية المفهوم والهدف وإن كان التكشيف يعطينا مرونة أكبر.

والآن هل يختلف تكشيف الوثائق عن تكشيف المفاهيم؟ وإذا كان يقصد بكشافات الوثائق تحديد الموضوعات لكل مواد الوثيقة Whole Items أو لأكبر الأجزاء .. ثم تجميع الوثائق بعد ذلك، فإن تكشيف المفاهيم يتصل بتحديد المفاهيم أو الأفكار داخل الوثائق أو تحليلها وتخليقها وإعادة تشكيل محتوياتها.

ومرة أخرى هل يمكن أن ننقل هذا التعريف ونترجمه بالنسبة لبيئة النص الكامل على الخط المباشر؟ إن ذلك يعتمد على شكل قاعدة البيانات الخاصة بالنص الكامل، فهناك بنوك المعلومات Information data bank وهناك قواعد بيانات وثائق النص الكامل، والأولى يمكن تكشيفها ككل، أما الثانية فقد تحتوى ـــ مثلاً ـــ على ثلاث موسوعات فى نفس قاعدة البيانات، وبالتالى فمن الأفضل لكفاءة الاسترجاع بالنسبة للمستفيد إذا تم تجميع المقالات المختلفة فى الموسوعات الثلاث بالنسبة لنفس الموضوع حتى لا يقوم المستفيد ببحث كل موسوعة على حدة.

ونخلص من هذا العرض الموجز إلى تعريف لكل من الكشاف والتكشيف فالكشاف هو دليل منهجى لمحتويات الكتب أو المقالات أو التقارير أو غيرها من المواد المطبوعة أو غير المطبوعة. ويتم تمثيل ناتج تحليل المواد أو الأفكار على هيئة مداخل مرتبة هجائياً أو زمنياً أو رقمياً أو غير ذلك من نظم الترتيب، هذا والمدخل يمكن أن يكون كلمة أو جملة مفتاحية أو رمز للدلالة على مكان المادة أو المواد المرتبطة بمفهوم معين.

ويمكن أيضاً أن نعرف الكشاف بأنه الدليل المنظم (المرشد المنظم) للمحتوى الفكرى والمكانى المادى لتسجيلات المعرفة، أى أن الكشاف مؤشر أى مرشد وبالتالى فهو لا يزودنا بالمعرفة أو المعلومات نفسها ولكنه يستخدم مجموعة التاجات Tags أو الواصفات والتى تدل على مصدر المعلومات التى يبحث عنها المستفيد.

أما التكشيف فهو عملية إعداد المداخل والروابط التى توصلنا للمعلومات فى مصادرها ويمكن التمييز بين التكشيف والمصطلح القريب منه وهو الفهرسة، على اعتبار أن الأخير يهتم بوصف الوحدات الكاملة، أما التكشيف فيهتم بتحليل الجزئيات التى تكون كليات أكبر .. أى أن الفارق هنا هو فى الدرجة وليس فى النوع.

٢- علاقة التكشيف بالاستخلاص

أنشطة التكشيف الموضوعى والاستخلاص ترتبط ببعضها ارتباطاً وثيقاً، ذلك لأن كلاً منهما يتضمن إعداد تمثيل للمادة الموضوعية فى الوثائق، وإذا كان القائم بالاستخلاص يكتب وصفاً سردياً أو ملخصاً للوثيقة، فإن الكشاف يصف محتوياتها باستخدام واحد أو أكثر من مصطلحات التكشيف.. وهذه عادة يتم اختيارها من قائمة مصطلحات محكومة، وإذا كان الغرض الأساسى للمستخلص هو بيان ما تتحدث عنه الوثيقة What is about فإن عدداً من مصطلحات الكشاف يمكن أن يخدم نفس الغرض (Lancaster, 1991,5) بمعنى أن قائمة المصطلحات فى مجال معين يمكن اعتبارها نوعاً من المستخلص المصغر Mini Abstract.

وخلاصة هذا كله أن التمييز بين التكشيف والاستخلاص أصبح عسيراً، ذلك لأن قائمة مصطلحات الكشاف يمكن طباعتها لتشكل مستخلصاً مصغراً، كما أن

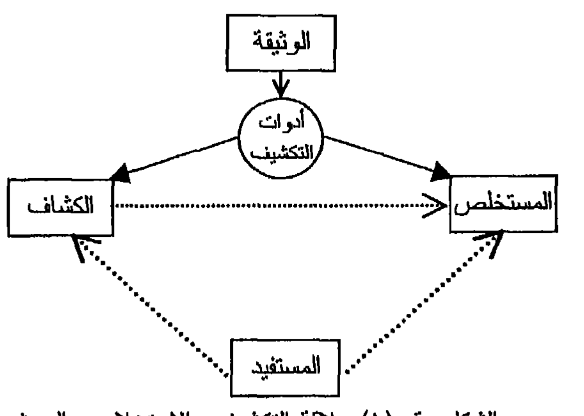
نص المستخلصات يمكن اختزانه فى النظام المعتمد على الحاسب بطريقة تسمح لنا بأداء البحث بمصطلحات من النص. وهذه المستخلصات يمكن استخدامها بدلاً من مصطلحات الكشاف. وذلك للوصول إلى المواد، أو أنها تكمل نقاط الإتاحة التى تقدمها لنا مصطلحات الكشاف.

ولعل ذلك أن يغير بعض الشيء من دور القائم بالاستخلاص، الذى يجب عليه الآن الاهتمام ليس بكتابة وصف واضح جيد لمحتويات الوثائق فحسب، ولكن عليه أن ينشيء تسجيلات ممثلة بدرجة كافية لأغراض الاسترجاع.

وإذا نظرنا إلى أنشطة التكشيف والاستخلاص بطريقة تكاملية، فإن طبيعة عملية التكشيف يمكن أن تتغير بعض الشيء، فالمكشف على سبيل المثال يمكن أن يركز على تعيين المصطلحات التى تكمل نقاط الإتاحة التى يقدمها المستخلص .. كما أن هذا التكامل بين التكشيف والاستخلاص لابد أن يكون مفهوماً بواسطة المستفيدين من قاعدة البيانات .. وإلا فإن مجموعة مصطلحات الكشاف وحدها ستؤدى إلى صورة مشوهة لمحتوى الوثيقة.

ثالثاً: أهمية التكشيف والاستخلاص فى استرجاع المعلومات:

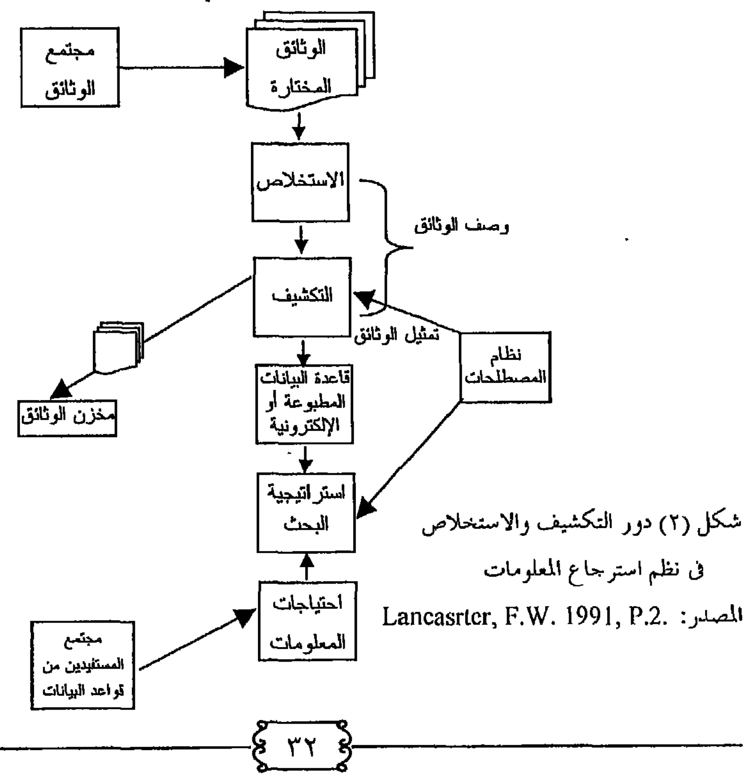
يضم نظام استرجاع المعلومات كلاً من عمليات التكشيف والاستخلاص والبحث، ذلك لأن الكشافات والمستخلصات تصبح بلا معنى إلا إذا استخدمت فى البحث، ويمكن توضيح العلاقات بالشكل التالى (Clevelend, D., 1990, 16) رقم (١):



الشكل رقم (١) علاقة التكشيف والاستخلاص بالبحث

ويلاحظ أن الخطوط المتقطعة تدل على الطريق الذى سيسلكه المستفيد، وقد يستخدم المستفيد بعض أدوات التكشيف (كالمكنز أو قوائم التصنيف) وإن كان ذلك غير مبين بالرسم أعلاه، هذا وقد يكون السبيل إلى المستخلصات نفسها عن طريق الكشاف، خصوصاً إذا كان للمستخلصات ترتيب مصنف وستنعكس هذه العلاقات فى رضى المستفيد من عدمه عن نظام الاسترجاع. فعلى سبيل المثال فالمستفيدون يهتمون بالزمن الذى يستغرقه تلبية سؤال عن المعلومات، فالتأخير الطويل الذى لا مبرر له، سوف يثبط من همة الباحث لاستخدام نظام معين سواء اعتمد على الحاسب أو على الكشاف اليدوى المطبوع.

أما بالنسبة لدور التكشيف والاستخلاص داخل الإطار العام لأنشطة استرجاع المعلومات فقد أوضحها لانكستر فى الشكل التالى رقم (٢). حيث أوضح أن منتج قاعدة البيانات يقوم باختيار الوثائق الحديثة التى تستجيب لمعايير معينة لإدخالها فى قاعدة البيانات .. وأهم هذه المعايير هو الموضوع الذى تعالجه القاعدة، وإن كانت أشكال الوثائق ولغتها ومصدرها ذات أهمية كذلك.



شكل (٢) دور التكشيف والاستخلاص فى نظم استرجاع المعلومات

المصدر: Lancasrter, F.W. 1991, P.2.

هذا ويجب وصف المواد الداخلة فى قاعدة البيانات بطرق مختلفة، لعل بدايتها الفهرسة الوصفية والتى (لا تظهر فى الشكل) لتحديد العناصر الببليوجرافية كالمؤلفين والعناوين والمصادر .. إلخ. أما إجراءات التكشيف فتحدد المادة الموضوعية، كما يستخدم الاستخلاص لتلخيص محتوى المواد .. والمصطلحات المستخدمة فى التكشيف يتم التقاطها عادة من أحد أشكال قوائم التحكم فى المصطلحات كالمكانز (والتى تظهر فى مربع نظام المصطلحات) بالشكل (٢).. وقد تكون مصطلحات حرة تؤخذ من الوثيقة نفسها .. وهذا الوصف من مختلف الجوانب ينشىء ما يسمى بتمثيل الوثيقة فى شكل مناسب لإدخالها فى قاعدة البيانات. أما الوثائق نفسها فتحفظ عادة فى شكل مختلف من قواعد البيانات (مخزن الوثائق) مثل رفوف المكتبة.

ويستخدم المستفيدون قاعدة البيانات لإرضاء حاجاتهم المعلوماتية، من أجل ذلك فيجب تحويل هذه الاحتياجات إلى شكل من أشكال "استراتيجية البحث" والتى قد تكون بسيطة كمصطلح واحد لاستشارة الفهرس المطبوع أو الفهرس البطاقى، أو قد تكون أكثر تعقيداً عند جمع مصطلحين أو أكثر لبحث قاعدة البيانات بواسطة النهايات الطرفية للحاسب الآلى.

رابعاً: مشكلة استرجاع المعلومات وإجراءات البحث:

يمكن أن تظهر مشكلة استرجاع المعلومات فى الأشكال (٣)، (٤)، (٥) كما يلى:

(أ) الاستدعاء والدقة: **Recall and Precision**

يهدف الطالب عند بحث قاعدة المعلومات العثور على مواد مفيدة تستجيب لبعض احتياجاته المعلوماتية، فى نفس الوقت الذى يحرص فيه على تجنب استرجاع مواد غير مفيدة له. والمواد المفيدة هذه يطلق عليها فى الإنتاج الفكرى المعلوماتى الصالحة أو

شكل (٣) مشكلة استرجاع مواد صالحة من قاعدة البيانات

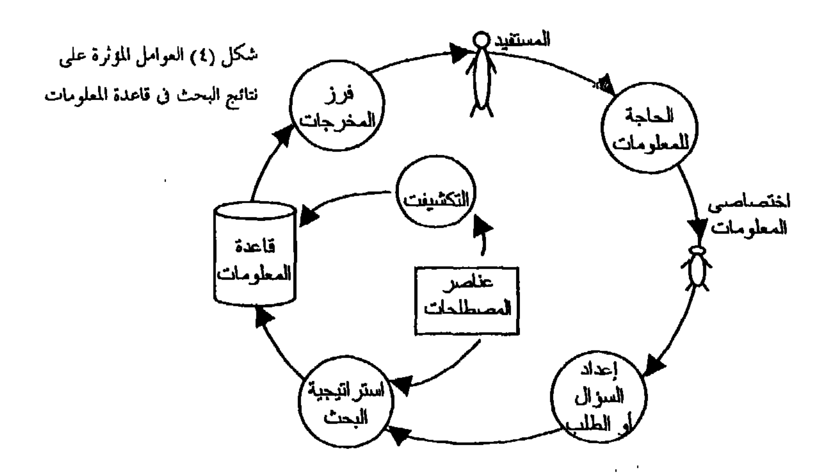
المصدر: Lancaster, 1991, P.3

المتعلقة Relevant or Pertinent ويعبر عن مشكلة استرجاع المعلومات بالرسم فى الشكل (٣) حيث يمثل المستطيل الكامل قاعدة البيانات والمواد التى تشملها، أما علامة (+) فتدل على ما سيجده الطالب مفيداً، والعلامة (-) تدل على المواد التى يراها هو غير مفيدة. والمشكلة هى فى كيفية استرجاع أكبر عدد ممكن من المواد المفيدة وأقل عدد ممكن من المواد غير المفيدة.

وإذا تفحصنا المستطيل الأصغر فى الرسم فسنجد أن الطالب وجد [٦] مواد مفيدة وكذلك [٥١] مادة غير مفيدة أى أن نسبة المواد المفيدة لإجمالى المواد التى تسلمها هو ٦/٥٧ أى حوالى ١٠% وهذه النسبة هى التى يطلق عليها نسبة الدقة Precision Ratio أما النسبة التى تدلنا على درجة وجود جميع المواد المفيدة فى قاعدة البيانات فتسمى نسبة الاستدعاء Recall ratio وفى هذه الحالة تكون ٦/١١ أى حوالى ٥٤% [لاحظ أن المواد (١١) هى جميع المواد المفيدة فى القاعدة].

وحتى يحسن الطالب من الاستدعاء فعليه مثلاً أن يبحث بطريقة أكثر اتساعاً أى لا يقتصر البحث على المستطيل الأصغر بل يتعداه إلى المستطيلين الأكبر. وعند قيامه بهذا البحث فإن الاستدعاء يرتفع مثلاً إلى [٨] أى أن نسبة الاستدعاء هى ٨/١١ (أى حوالى ٧٣%) ولكن ولكن نسبة الدقة ستنخفض لتصبح ٨/١١٢ ى حوالى ٧% وهذه هى المشكلة الرئيسية لأن تحسين الاستدعاء سيؤدى عادة إلى تدهور فى الدقة والعكس صحيح.

ويلاحظ أن سلوك الباحث يدلنا على استعداده للنظر فى [٥٧] مستخلص مثلاً من قاعدة البيانات ليعثر على [٦] مواد مفيدة، ولكنه عادة غير مستعد للنظر فى [٥٧٠] مستخلص ليعثر على [٦٠] مادة مفيدة، أى أننا مع قواعد المعلومات الكبيرة، فهناك صعوبة متزايدة لتحقيق مستوى مقبول منه الاستدعاء عند مستوى محتمل الدقة، ويستخدم العالم لانكستر (Lancaster, 1991, P.4) مصطلح الاستدعاء Recall للدلالة على القدرة على استدعاء المواد المفيدة ومصطلح الدقة Precision للدلالة على القدرة على تجنب المواد غير المفيدة.



شكل (٤) العوامل المؤثرة على نتائج البحث فى قاعدة المعلومات

(ب) بعض العوامل التى تتحكم فى نتائج بحث قاعدة البيانات:

هناك عوامل عديدة تحدد مدى نجاح البحث فى قاعدة المعلومات، وتتضمن تلك العوامل مدى تغطية القاعدة، سياسية التكشيف، ممارسة التكشيف، سياسة وممارسة الاستخلاص، نوعية المصطلحات المستخدمة فى الكشاف ونوعية استراتيجية البحث ...إلخ.

فالتكشيف الجيد هو الذى يتيح للطالب استرجاع المواد التى تستجيب لاحتياجاته، ويدلنا الشكل (٤)، على العوامل التى تتحكم وتؤثر فى أداء نظام استرجاع المعلومات، فالطالب عندما يحتاج لمعلومات معينة يقوم بالتشاور مع اختصاصى المعلومات، ونتيجة ذلك يتحدد السؤال أو الطلب وبناءً على ذلك يقوم اختصاصى المعلومات بإعداد استراتيجية بحث باستخدام المصطلحات الكشفية وكلمات النص أو بعض التوليفات منهما. ثم يقوم اختصاصى المعلومات بمضاهاة هذه الاستراتيجية على قاعدة المعلومات. وكنتيجة للبحث يتم استرجاع بعض المواد، وهذه بدورها يتم فرزها بواسطة الباحث لاستبعاد ما يعتقده الطالب أنها مواد غير مفيدة، وأخيراً هناك مجموعة من الوثائق أو المراجع تسلم للمستفيد.

وواضح ضرورة تمثيل السؤال أو الطلب (الشكل رقم٤) للاحتياجات الفعلية للمستفيد، أما العامل الثانى فهو نوعية استراتيجية البحث، وتدخل هنا عوامل الخبرة والذكاء والإبداع، ومع ذلك فالمصطلحات المستخدمة فى قاعدة المعلومات لها أهميتها

الأساسية. فإذا كانت المصطلحات المقيدة هى المستخدمة، فلا يستطيع الباحث إلا أن يكون داخل هذه الدائرة، وإن كان من الممكن الوصول إلى تخصيص إضافى عن طريق استخدام كلمات النص. والمشكلة هنا مرة أخرى هى إيجاد التوازن بين الاستدعاء والدقة، والحاجة عادة هى الوصول إلى أعلى معدل فى الاستدعاء مع مستوى مقبول من الدقة.

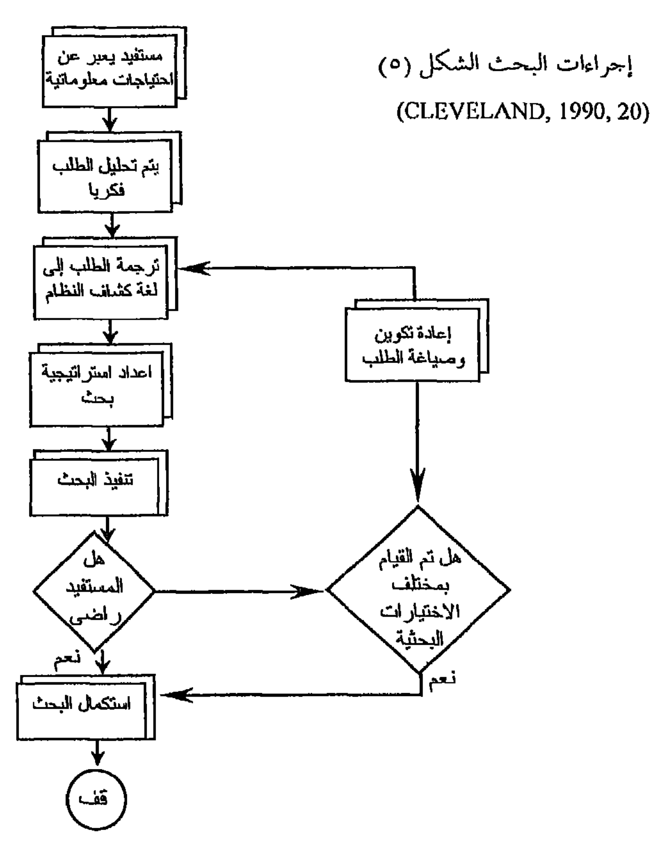
وإذا كان الأداء والمخرجات تعتمد على عوامل عديدة، فهناك عاملان لهما الأولوية هما:

* مدى فهم اختصاصى المعلومات لما يحتاجه المستفيد فعلاً.

* مدى كفاءة تكشيف الوثائق المختزنة فى قاعدة المعلومات ومدى تعبيره عن المحتوى الفعلى للوثائق وهذه الكفاءة فى التكشيف تتضمن عوامل عديدة أيضاً، أهمها عدم الفشل فى التعرف على الموضوع ذى الأهمية للمستفيدين من الوثيقة وكذلك عدم الفشل فى استخدام المصطلح الأكثر تخصيصاً وتحديداً وتعبيراً عن الموضوع. وما يترتب على ذلك من نجاح كل من الاستدعاء والدقة (Lancaster, 1991, P. 74-76).

وأخيرا فيرى كليفيلند أن هناك نظاماً معيارياً لهذه الإجراءات البحثية ويتمثل ذلك فى الشكل (٥) التالى، حيث يسأل الطالب سؤالاً يعبر به عن حاجته المعلوماتية، ويترجم السؤال إلى نفس الكلمات أو الرموز المستخدمة لتمثيل الوثيقة .. وبمعنى آخر يتم تحليل السؤال وتكشيفه حتى تتطابق لغة الطالب مع اللغة الموجودة بالتسجيلات .. ثم تتم استراتيجية البحث ثم القيام بالبحث سواء بالطريقة اليدوية أو بالحاسب الآلى.

هذا وتعتبر منطقية البحث وسيلة تحديد توليفات المصطلحات المستخدمة فى الاسترجاع الناجح عن طريق المضاهاة Matching ، وتستخدم منطقية البحث البولينى فى معظم نظم الاسترجاع، حيث يتم ربط المصطلحات من لغات التكشيف المحكومة أو الطبيعية أو كلاهما، ويستخدم المنطق لربط المصطلحات التى تصف المفاهيم الموجودة فى البحث، وقد تصل هذه المصطلحات إلى عشرين مصطلح بحثى.



إجراءات البحث الشكل (٥)
(CLEVELAND, 1990, 20)

وتسمح منطقية البحث باستخدام جميع المترادفات والمصطلحات ذات العلاقة، كما تحدد توليفات مصطلحات البحث المقبولة وغير المقبولة، وتتطلب استراتيجيات البحث أن تكون أكثر تعقيداً مع مصطلحات اللغة الطبيعية، وذلك لاحتوائها على جميع اختلافات نطق وهجاء الكلمات والمترادفات القريبة.



الفصل الثانى

الكشافات والتكشيف

دراسة فى أنواع الكشافات ومستويات التكشيف وخطواته

مقدمة

يتناول هذا الفصل مستويات التكشيف للكلمات والأسماء والكتب والدوريات ونظم استرجاع المعلومات، كما تتناول الدراسة أيضاً أنواع الكشافات الخاصة بالمؤلفين والكشافات الموضوعية الهجائية والمصنفة وكشافات الترابط وكشافات العناوين الدوارة والكشافات الوجهية والمتسلسلة ثم نظم التحليل الكشفى للفهرس المصنف وكشاف الخيوط الرابطة String وأخيراً كشافات الاستشهادات المرجعية أما الجزء الثانى من الدراسة فيتناول عملية التكشيف وخصائصها بما يتضمنه ذلك من تعريف بمشكلات الاحتياجات المعلوماتية للمستفيدين، وببعض القواعد العامة فى التكشيف ثم خطوات عملية التكشيف نفسها بما فى ذلك قياسات الاستدعاء Recall والدقة Precision .

أولاً: مستويات التكشيف

هناك مستويات عديدة للتكشيف يمكن أن يشار إليها فيما يلى:

(أ) كشافات الكلمات والأسماء.

(ب) كشافات الكتب.

(ج) كشافات الدوريات.

(د) كشافات نظم استرجاع المعلومات.

وداخل هذه المستويات هناك الأنواع أو الأشكال التالية:

١- كشافات المؤلفين

٢- الكشافات الموضوعية الهجائية.

٣- الكشافات المصنفة.

٤- كشافات الترابط.

٥- كشافات العناوين الدوارة Permuted

٦- كشافات الأوجه (الكشافات الوجهية)

٧- الكشافات المتسلسلة. Chain Indexes

٨- نظم التحليل الكشفى للفهرس المصنف.

٩- كشاف الخيوط الرابطة. String

١٠- كشافات الاستشهادات المرجعية.

(أ) كشافات الكلمات والأسماء:

يمكن أن تكشف تسجيلات المعرفة طبقاً للكلمات الفعلية الموجودة فى النص وفى هذه الحالة نحن لا نستخدم قائمة استشهاد معينة ويمكن أن يطلق على كشافات الكلمات والأسماء فى بعض الأحيان كشافات التواتر Concordances (كشافات النصوص) وكشافات النصوص/ هى كشافات للأسماء أو الكلمات التى يستخدمها المؤلف للتعبير عن المعلومات فى أثناء كتابته للمخطوط.

وهذه الكشافات ذات أهمية للمستفيدين خصوصاً لعلماء اللغة عندما يكون المصطلح أو الكلمة نفسها ـ داخل النص ـ هى التى تحدد الموضوع ومكانه ولكن هناك حدوداً أو صعوبات بالنسبة للإفادة من هذه الأنواع من الكشافات وإن كانت ذات أهمية بالنسبة للكتب الدينية.

وواضح أن هذا النوع من التكشيف سهل للمكشف لأنه سوف لا يستخدم أحكامه الشخصية كما يمكن أن يتم بواسطة أشخاص ذوى تدريب قليل كما يمكن أن يتم بواسطة الآلات أى أنه يمكن أن يتم بسرعة وتكاليف قليلة.

ولكن من العيوب الرئيسية لهذه الكشافات أنه يعقد عملية البحث بطرق مختلفة نظراً لانتشار المعنى فى مصطلحات مترادفة عديدة كما يتجاهل أخطاء الكتابة، كما أنه لا يعكس أى علاقات بين المصطلحات خصوصاً علاقة العام إلى الخاص، وهى علاقة لابد وأن توجد فى عملية التكشيف، كما أن الباحث لابد وأن يلجأ من مصطلح إلى آخر حتى يمكن أن يحيط بالمداخل الممكنة لموضوع اهتمامه.

ولما كانت إحدى مميزات الكشاف الجيد هو إمكانية التركيز السريع على مصطلحات المداخل التى تعبر عن احتياجات المستفيدين فضلاً عن ربط لغة التكشيف بطريقة تفكيرهم فإن كشافات النصوص وهى كشافات غير محكومة

تستهلك الوقت وتعتبر عبئاً على الباحث، أى أن السرعة والاقتصاد الذى يحققه إنشاء مثل هذه الكشافات يكون على حساب وقت وجهد المستفيد.

وحتى فى حالة استخدام الحاسب الآلى فلابد أن يعاود الباحث الأسئلة واستخدام مصطلحات بديلة، ذلك لأنه ليس هناك ضبط للمصطلحات، والحاسب الآلى فى هذه الحالة يتيح للمستفيد أن يجد الأشياء غير الصحيحة بطريقة أسرع (Cleveland, D., 1990, 28.).

ومع ذلك فقد أثبت هذا النوع من الكشافات أهم يته بالنسبة لكشاف أرقام براءات الاختراعات وكشافات المعادلات أو المركبات الكيميائية ويستشير المستفيد الكشاف طبقاً للمعادلة Formula أو العناصر الكيميائية حيث تعتبر هذه الطريقة أسهل من البحث بأسماء المركبات نفسها.

وخلاصة ذلك أننا لا نستطيع إهمال كشافات الكلمات بل هى صالحة فى بعض الأحيان.

(ب) كشافات الكتب:

تمثل كشافات الكتب قطاعاً كبيراً من الكشافات التى يعرفها الجمهور القارىء أكثر من غيرها وفى الواقع فعندما يذكر كشاف الكلمات فمعظم الأشخاص يفكرون فى كشاف الكتاب، ذلك لأنه يحتوى على قوائم بالكلمات مرتبة هجائياً عادة فى نهاية الكتاب ويحدد مع الكلمات مكان الصفحة والأسماء المرتبطة بكل كلمة، ومع ذلك فهذا الكشاف ليس بديلاً للمعلومات الموجودة بالكتاب ولكنه مجرد مؤشر للمعلومات الموجودة فيه.

وهذه الكشافات هامة فى معظم الموضوعات باستثناء الروايات والقصص وكتب الرحلات التى سيكون لها كشافات بأسماء الأماكن والأحداث التاريخية والفنادق فضلاً عن أماكن الترويح، وكذلك الحال بالنسبة لكتب التراجم مع التركيز على أسماء الأماكن والأحداث التاريخية والناس.

أما الكتب البحثية فتتطلب كشافات كاملة ومعقدة بحيث تعكس هذه الكشافات العلاقات المختلفة والأفكار المجردة فضلاً عن البيانات الأخرى.

ومن أمثلة كشافات الكتب الهامة والمتميزة كشافات الموسوعة البريطانية خصوصاً المايكروبيديا Micropeadia فى طبعتها الخامسة عشرة.

(ج) كشافات الدوريات:

فى منتصف عام ١٦٠٠ بدأت أول دورية ثم استمر نشر الدوريات بمعدل تضاعفى كل ٥٠ سنة وظهرت حيرة وعجز الباحثين بالنسبة لنمو المعلومات التضاعفى وما نتج عنه من صعوبة اللحاق بما ينشر، وبالتالى ظهرت الحاجة إلى أساليب فعالة للضبط البيليوجرافى وللوصول للإنتاج الفكرى للدوريات حسب الموضوع مما أدى إلى نتيجة طبيعية وهو الكشاف الدورى لمحتويات الدوريات.

وكشافات الدوريات تعتمد على نفس المبادئ، ولها نفس الأهداف الخاصة بكشافات الكتب وإن كان نطاق كشافات الدوريات أعرض وله مشاكله المتميزة، وعلى سبيل المثال فكشاف الكتاب هو عملية يقوم بها شخص واحد عادة نظراً لأنه يركز على موضوع عام من البداية للنهاية.

أما كشافات الدوريات فهى مشروعات مفتوحة النهايات ويتم إعدادها بواسطة فريق من الأفراد، حيث يقومون عادة بتغطية عدة سنوات فى بعض الأحيان، ومع اختلاف فى موضوعات الدوريات.

ولما كان الانتظام فى عملية التكشيف ضرورياً فإن كشاف الدورية يجب أن يعمل على توحيد أو تنظيم أو تقليل الاختلافات التحريرية والتصنيفية الموجودة فى مختلف الدوريات التى يتم تكشيفها.

كما أن هناك بعض القرارات التى يجب أن تتخذ بالنسبة للإعلانات مثلاً التى توجد بالدورية (هل يشملها التكشيف من عدمه). كما أن المصطلحات قد تتغير مع الزمن وبالتالى يجب أن تتغير الكشافات لتواكب هذا التغيير ومن أمثلة كشافات الدوريات ما يلى:

1- Readers Guide to Periodical Literature. New York: Wilson 1905. V.1-

2- New York Times index. New York: Times 1913 V.1-

3- Library Literature 1921/32. New York: Wilson 1934-

وهناك شكلان من كشافات الدوريات الأول: الكشافات الفردية التى تعد لدورية محددة، وثانيهما الكشافات التى تعد لأكثر من دورية.

(د) كشافات نظم استرجاع المعلومات:

يمكن تعريف استرجاع المعلومات بأنه الاستدعاء الانتقائى والمنهجى للمعلومات المختزنة بطريقة منطقية.

ونظام استرجاع المعلومات فى معناه العام لا يتطلب وجود التكنولوجيا بالضرورة، وإن كان الاستخدام العام خصوصاً فى الوقت الحاضر يتضمن الحاسبات الآلية.

وما ينبغى الإشارة إليه أن المكتبة التقليدية مهما كان تعريفها هى نظام استرجاع معلومات، وكذلك دليل التليفون ولكن الاستخدام المعاصر يتضمن وجود التكنولوجيا خصوصاً الحاسب الآلى. وهناك بعض الوظائف التى يجب أن يقوم بها نظام استرجاع المعلومات وهى:

(أ) الحصول على الوثائق الضرورية والمناسبة.

(ب) إعداد وتمثيل محتوى هذه الوثائق.

(ج) تكويد مؤشرات المحتوى لسهولة التطويع.

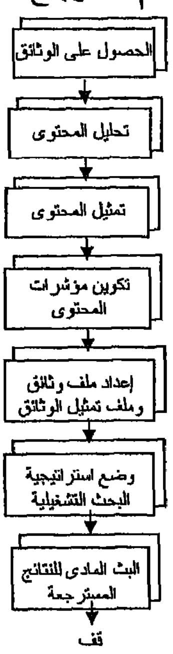
(د) الاختزان المنظم لهذه الوثائق ومؤشراتها فى ملفات مستقلة.

(هـ) تطوير استراتيجيات البحث.

(و) البث المادى لنتائج الاسترجاع.

ويلاحظ أن الإجراء الذى يتعلق بتحديد وتمثيل المجموعة للمستفيد يقع فى مركز النظام وهو الكشاف وذلك طبقاً للوظائف الموضحة فى الرسم التالى لنظام استرجاع المعلومات.

وظائف نظام استرجاع المعلومات



ثانياً: أنواع الكشافات

١- كشافات المؤلفين:

ترتيب المواد هنا تحت أسماء مؤلفيها سواء أكانوا أفراداً أم هيئات أم جامعات أو غير ذلك، وقد يستخدم المؤلفون بطريقة غير مباشر للتعرف على الموضوع، ذلك لأن المتخصصين فى مجال موضوعى معين يعرفون جيداً أسماء المشهورين فى المجال من خلال هؤلاء المؤلفين، وقد أوضح كليفيلند (Cleveland, 1990, 40) أن المؤلفين فى بعض أنواع نظم الاسترجاع يعتبرون مؤشرات للمحتوى الموضوعى فى مجموعة من الوثائق وهناك مشكلات فى إعداد هذه الكشافات من بينها عدد الأسماء التى يمكن تخصيصها لكل مدخل عندما يكون للوثيقة مؤلفون متعددون، وكذلك مشكلة الترتيب الهجائى، استخدام الاسم الكامل أو الحروف الأولى ومن أمثلة هذه الكشافات Book Review Digest. New York: HW Wilson وهو ملخص وكشاف لمراجعات بعض الكتب المختارة.

٢- الكشاف الموضوعى الهجائى: Alphabetical Subject index

يغطى مصطلح الكشاف الهجائى عدداً من أنواع الكشافات، ويتم الترتيب هجائياً .. ولكن الكشاف يمكن أن يتبع الترتيب المصنف، ولكن الترتيب المصنف يحتاج إلى مسلك هجائى كملحق، وأهم مشكلات الكشاف الهجائى هى المترادفات وتبعثر المداخل المرتبطة ببعضها وكذلك اختيار طريقة الترتيب (كلمة بكلمة أو حرف بحرف) ومن أمثلتها: Applied Science and Technology index. New York: H. W Wilson.

وإذا كانت المكتبات تقوم بعمل الإحالات من المصطلح الغير مستخدم إلى المصطلح المستخدم إلا أن ذلك يسبب تعقيداً للكشاف وزيادة حجمه، وإذا كان الترتيب حسب الأحرف يعتبر أسهل، فالترتيب حسب الكلمات يعتبر أكثر صعوبة، أما بالنسبة للأرقام فيتم التعامل معها حسب نطقها ونفس الطريقة بالنسبة للمختصرات وعلى كل حال فالترتيب الهجائى ليس عملية مباشرة كما قد يظن البعض، ولكنها تحتاج للقواعد المقننة.

٣- الكشافات المصنفة:

الكشاف المصنف يرتب هرمياً حسب الموضوعات المتعلقة ببعضها مبتدئاً بالموضوعات الأكثر عمومية، وننزل هرمياً إلى الموضوعات الأكثر خصوصية، وواضح أن مثل هذه الخطة لابد أن تكون موجودة قبل القيام بعملية التكشيف، وذلك لأن كل مفهوم له مكانه المحدد فى التسلسل الهرمى.

والكشافات المصنفة لها مزاياها وعيوبها، فالمزايا الواضحة هى أنها تعتبر فكرياً معاونة للبحث لأنها تتضمن مفاهيم من العام إلى الخاص خصوصاً والمستفيدون متعودون على الترتيب المنطقى للمعرفة وبالتالى فالكشافات المصنفة تجعل البحث يسيراً إذا أراد الباحث القيام بالبحوث على المستويات العامة. وعندما يحدد الباحث مدخلاً معيناً فهو يستطيع التعرف على المواد أو المفاهيم القريبة منه أعلاه وأدناه.

أما العيب الرئيسى فى الكشاف المصنف فهو ضرورة وجود قائمة هجائية كملحق ثانوى وبالتالى فالبحث يتطلب القيام بخطوتين أولهما استشارة القائمة الهجائية ثم الذهاب إلى الكشاف المصنف.

والإنتاج الفكرى ملىء بالمناقشات حول أفضلية الكشاف الهجائى، على الكشاف المصنف أو العكس، والحقيقة أن كلا منهما له مزاياه وعيوبه. ومن أمثلة الكشافات المصنفة كشاف الرسائل المقدمة للجامعات البريطانية.

Index to theses accepted for higher degree in the Universities of Great Britain and Ireland 1950/51 V.1 London: Aslib 1953-

وفيما يلى أنواع الفهارس الموضوعية الثلاثة والمقارنة بينها (محمد فتحى عبد الهادى، ١٩٧٧).

(أ) من أمثلة الفهرس الهجائى المصنف ما يلى:

الاجتماع (علم)

الاجتماع (علم) ــ البيئة والمجتمع

الاجتماع (علم) ــ التغير الاجتماعى

المكتبات (علم)

المكتبات (علم) ــ العمليات المكتبية

المكتبات (علم) ــ العمليات المكتبية ــ الإعارة

(ب) ومن أمثلة الفهرس المصنف لنفس الموضوعات السابقة ما يلى:

٠٢٠ علم المكتبات

٠٢٥ العمليات المكتبية

٠٢٥,٦ الإعارة

٣٠١ علم الاجتماع

٣٠١,٢٤ التغير الاجتماعى

٣٠١,٣ البيئة والمجتمع

(ج) ومن أمثلة الفهرس الموضوعى الهجائى لنفس الموضوعات:

الاجتماع (علم)

انظر أيضاً: البيئة والمجتمع. التغير الاجتماعى

الإعارة

البيئة والمجتمع

التغير الاجتماعى

الخدمات المكتبية

انظر أيضاً: الإعارة

العمليات المكتبية

انظر أيضاً: الإعارة. التجليد والترميم. الخدمات المكتبية

المكتبات (علم)

انظر أيضاً: العمليات المكتبية

ومع ذلك فالفهرس الموضوعى الهجائى له عيوبه منها ما يلى:

١- أنه يفتقد التجمعات المنطقية الموجودة فى الفهرس المصنف.

٢- لا يفيد كثيراً فى التعاون الدولى مثل الفهرس المصنف حيث يعتبر الرمز التصنيفى كلغة عالمية.

٣- ليس من السهل طبع أجزاء مستقلة منه مثل الفهرس المصنف ومن الصعب الاعتماد عليه فى إعداد الببليوجرافيات.

٤- هناك بعض الضعف فى بناء المفاهيم الخاصة برؤوس الموضوعات.

كما يصعب المحافظة على حداثته لأن التغيير فى المصطلحات يؤثر فى مداخل كثيرة.

٤- كشافات الترابط: Coordinate indexes

تنشأ كشافات الترابط هذه عن طريق جمع اثنين أو أكثر من مصطلحات التكشيف المنفردة لإعداد فئة أو موضوع جديد New Class وتنقسم كشافات الترابط إلى نوعين رئيسين أولهما الكشافات سابقة الترابط precoordinate indexes وهنا تتم عملية الترابط فى مرحلة التكشيف .. وهذا النوع من التكشيف ضرورى بالنسبة للكشافات المطبوعة التقليدية، نظراً لأن الصفحة المطبوعة تعتبر المنتج النهائى.

وثانيهما الكشافات لاحقة الترابط Post coordinate indexes وتسمى أحياناً بكشافات التطويع (Manipultive indexes) وهنا يتم الترابط بواسطة المستفيد فى مرحلة البحث، وليس بواسطة المكشف فى مرحلة التكشيف .. أى أن المستفيد يقوم باستراتيجية بحث عن طريق جمع المصطلحات بأدوات بوليان Boolean Operators للتعبير عن الاحتياجات المعلوماتية على قدر المستطاع.

ولقد قام الكيميائى باتن Batten فى الأربعينيات بتطبيق هذا الأسلوب فى بريطانيا وذلك لتكشيف الوثائق الكيميائية، حيث خصص بطاقة لكل مصطلح من مصطلحاته الموضوعية كما حدد أماكن ثابتة على البطاقة لوثائق معينة .. وعندما يخصص مصطلح معين لوثيقة معينة أثناء عملية التكشيف، فإن بطاقة المصطلح تسحب ثم يثقب عليها ثقب فى المكان الذى يمثل الوثيقة.

أما كالفن مورز Mooers وهو أحد الرواد فى استرجاع المعلومات فقد قام عام ١٩٤٧ باكتشاف أسلوب جديد عن طريق تمثيل الوثيقة ببطاقة (يسمى الأسلوب السابق لباتون بنظام الموضوع بينما يسمى أسلوب مورز بنظام الوثيقة)، فالمصطلحات المستخدمة فى تكشيف الوثائق يتم تثقيبها على أطراف البطاقات بطريقة التكويد العشوائى، وعند مرحلة البحث تدخل الإبرة فى مجموعة الوثائق، والبطاقات ذات التجمعات البولينية المطلوبة هى التى تسقط نظراً لوجود الثقوب عند الحافة، وهذه الطريقة تسمى أيضاً Zato Coding .

أما المدخل الهام الثالث فهو ما قام به مورتيمر تاوبه Taube عام ١٩٥١ فيما يسمى بنظام المصطلح الواحد وهو تعديل لأسلوب باتن Batten السابق الإشارة إليه حيث لم تعد الأماكن على البطاقات محددة لوثائق معينة، إذ قام بوضع أرقام تمثل الوثائق على البطاقات ولكن ترتيبها يتم حسب الرقم الأخير (جميع الأرقام التى تنتهى بالصفر تكون مع بعضها فى ترتيب تنازلى) .... وكان يأخذ المصطلح الموحد Uniterm مباشرة من النص، ولم يهتم بضبط المصطلحات، ولكن تبين له فيما بعد أن جمع بعض المصطلحات هذه للرد على استفسار معين، يشكل صعوبات دلالية Semantic فى عملية الاسترجاع، وقد دخلت أدوات ضبط المصطلحات شيئاً فشيئاً فى هذا النظام فضلاً عن إدخال مصطلحات متعددة الكلمات Multi Word Terms .

وإذا كانت الأساليب الثلاثة السابقة تسمى التكشيف لاحق الترابط، فإن كشافات الكتب والكشافات على هيئة كتاب تعتبر أمثلة للكشافات سابقة الترابط.

ويجب الإشارة هنا إلى أن الكشافات لاحقة الترابط أصبحت مفيدة وشائعة كأساس لمعظم نظم الاسترجاع على الخط المباشر.

وتتميز جميع نظم التكشيف اللاحق الترابط بالخصائص الثلاثة التالية:

(أ) جميع مداخل النظام غير محددة Not Specific، وهناك عدد كبير نسبياً من الوثائق فى كل موضوع، وإذا حاول الباحث استخدام الكشاف مثل الكشاف التقليدى فسيقوم بالبحث فى عدد كبير جداً من المداخل تحت كل موضوع حتى يستطيع أن يفرق بين الوثائق المتعلقة والوثائق الأقل تعلقاً بالبحث الذى يريده.

(ب) عدد مداخل نظام التكشيف اللاحق الترابط أكبر عادة من مداخل نظام التكشيف السابق الترابط، على الرغم من أن عدد المداخل فى نظام التكشيف السابق الترابط يعتمد على وجود المراجع والمداخل المتعددة.

(ج) عدد الموضوعات المختلفة فى الكشاف صغير نسبيًا نظراً لأنه ــ كالحال فى التصنيف ــ فإن النظام التركيبى يحتاج إلى فئات أو موضوعات أقل من النظام الحصرى المماثل.

٥- كشافات العناوين الدوارة: Permuted

تستعين الكشافات الدوارة بالحاسب الآلى لتدوير المصطلحات حتى يظهر كل مصطلح كمصطلح أول فى قائمة الترتيب الهجائى. وهذه مثل كشافات KWIC (Key Word in Context) أو كووك Kwoc (Key word out of Context) والتى تعتبر كشافات سابقة الترابط ومن مزايا هذه الكشافات أنه يمكن إنجازها بسرعة بتكاليف زهيدة ويتم عملها كاملاً بواسطة الحاسب الآلى، أى أنه ليس هناك تأخير بسبب المكشفين أو المحررين.

ويمكن الإشارة إلى بعض تلك الكشافات كما يلى:

٥-١ كشاف الكلمات الدالة فى السياق KWIC

ويسمى أحياناً كشاف التباديل Permutation index لأنه يعتمد على الكلمات المفتاحية أو الهامة فى عنوان ما، وذلك بعد استبعاد بعض الكلمات غير ذات الدلالة، وترتيب الكلمات المفتاحية ترتيباً هجائياً، ويأتى بعدها بقية كلمات العنوان، وهكذا يتكرر العنوان تبعاً لعدد الكلمات الدالة (التى تعتبر الرأس) والسياق ثم الكود (الذى قد يكون رقم مسلسل أو رقم المجلد متبوعاً بالصفحة)

ويستطيع الحاسب الآلى إنتاج مثل هذه الكشافات بسهولة .. وعلى سبيل المثال فإن نشرة العناوين الكيمائية Chemical Titles تقوم بدور إعلام الباحثين بسرعة بالبحوث الكيميائية والتى تصدر لها مستخلصات بعد ذلك فى دورية المستخلصات Chemical Abstracts .

٥-٢ كشاف الكلمات الدالة من خارج السياق:

Key Word out of Context (KWOC)

وفى هذا الكشاف تؤخذ الكلمات التى يتم بناء عليها ترتيب العناوين هجائياً فى الـ KWIC من مكانها فى العنوان، حيث توضع فى الهامش الأيسر (بالنسبة للعناوين الإنجليزية) بينما يرد العنوان كاملاً مشتملاً على كلمة الترتيب نفسها على يمين الهامش الأيسر ومشكلة هذا الكشاف هى كيفية معالجة المصطلحات المركبة أى المكونة من أكثر من كلمة واحدة.

٥-٣ كشاف الكلمات المفاحية المضافة إلى السياق:

Keyword Augmented to Context (KWAC)

وينشأ هذا الكشاف لمواجهة مشكلة كشاف الكلمات المفتاحية فى السياق الذى يعتمد على عنوان المقال، فقد لا يعبر هذا العنوان بصدق ودقة عن محتوى المقال .. أو على الأصح أن المؤلف لم يقم باختيار كلمات عنوانه لتصلح كمداخل كشفية، وبالتالى فقد جاء التفكير لإضافة بعض المصطلحات التى تدعم القوة الدلالية للعناوين.

ومع ذلك فإن كشاف العنوان البرميوتى Permuted Title Index ليس مفيداً مثل الكشاف الموضوعى المعتمد على تحليل الأجزاء الرئيسية للوثيقة. كما أن أهم عيوب تكشيف العناوين هى:

(أ) قد لا تعكس العناوين بدقة المحتوى الموضوعى.

(ب) العدد المحدود من المصطلحات فى العناوين يحد من التعبير عن الموضوع.

(ج) معظم هذه الكشافات لا تجذب الباحث وعسيرة فى البحث.

(د) عدم وجود ضبط للمصطلحات قد يؤدى إلى زيادة استرجاع الوثائق غير الصالحة.

(هـ) انتشار المترادفات والمصطلحات العامة مما يؤدى إلى فقدان مداخل هامة للباحث.

ومن الغريب أن هناك اهتماماً كبيراً فى الإنتاج الفكرى المعاصر بخصائص الكشافات المطبوعة سابقة الترابط فى عصر يعتبر فيه البحث على الخط المباشر لقواعد البيانات أمراً شائعاً متعارفاً عليه .. ولعل السبب فى ذلك يرجع إلى أن التجهيز بالحاسب الآلى يزودونا بوسائل كافية لتوليد المداخل المتعددة بطريقة منهجية من مدخلات واحدة يعدها القائم بالتكشيف Indexer .

ولعلنا نذكر أن هناك واحداً من المداخل الأكثر تعقيداً فى التكشيف السابق الترابط قام به فارادان Jason Farradane منذ أكثر من ثلاثين عاماً وأُطلق عليه اسم التكشيف العلاقى Relational Indexing وهى طريقة تتضمن استخدام مؤشرات الدور Role Indicators (أو عامل الدور Operators) وذلك للتعبير الدقيق عن العلاقات بين المصطلحات (Brookes, 1986) .

٦- **الكشافات الوجهيه والمكنز الوجهى:** Thesauro Facet

النظام الوجهى هو نظام ترابط قبلى وهو شكل من أشكال التصنيف التركيبى ويطلق عليه نظام تحليلى تركيبى والنظام الوجهى هذا يختلف عن نظم التصنيف الحصرية من حيث أن مصطلحاته تستخدم فقط كوحدات بناء وقد كان

العالم رانجاناثان هو أول من استخدم هذا المصطلح وأدخل فكرة النظام التصنيفى الوجهى.

وهو فى أساسه شبيه بالنظام المستخدم فى علم الحيوان حيث نبدأ بالقسم ثم بالأقسام الفرعية ثم بالأنواع ويتم هذا التسلسل عن طريق تحديد الخصائص التى تختلف بين كل مجموعة وأخرى.

والأساس الفلسفى وراء هذا النظام هى أن ورقة البحث هى خلق جديد حيث ينظر المؤلف إلى الموضوع بطريقة مختلفة، عند تسجيل أفكاره واكتشافاته الجديدة. ومعنى ذلك أن محتويات الوثيقة الجديدة لا يمكن أن يعبر عنها بدقة نظام التصنيف المبنى على المعلومات القديمة ومعنى ذلك أننا فى حاجة إلى خطة تصنيفية ديناميكية حتى تعكس الطبيعة الديناميكية للمعرفة ذاتها، ففى النظام الوجهى نقوم بوضع المصطلحات التى تمثل المفاهيم المعلوماتية فى الورقة البحثية الجديدة، أى أن الوجه هو قائمة للكلمات بحيث يكون كل مصطلح ذا علاقة دقيقة بالموضوع الذى يعتبر جزءاً منه.

وقد اقترح رانجاناثان مجموعة من الأفكار الأساسية الشاملة كما يلى (PMEST):

الشخصية – Personality
المادة – Matter
الطاقة – Energy
مكان – Space
زمان – Time

ويمكن أن نعتبر هذا النظام بشكل أو آخر شبيها بالمصطلحات الفردية فى تكشيف الترابط، ولكن ترتيب الوجوه يعتمد على احتياجات المستفيد، وعلى الطريقة التى سيتعامل بها المستفيدون مع النظام، كما يستخدم رانجاناثان الحروف كأكواد للدلالة على المصطلحات الوجهية وذلك لتوفير المكان وللمعاونة فى توضيح الوجوه المختلفة خصوصاً إذا تم ميكنة نظام الاسترجاع.

المكنز الوجهى Thesaurofacet :

يعتمد هذا المكنز والذى طبق بنجاح فى انجلترا، على ربط ودمج التصنيف الوجهى Faceted Classification بالمكنز والذى يستخدم أيضاً ككشاف موضوعى وقد ظهر المكنز الوجهى لأول مرة عام ١٩٦٧ كتطبيق على مجال الهندسة والموضوعات المرتبطة بها.

شكل وتكوين المكنز الوجهى:

فى المكنز الوجهى يحل المكنز محل الكشاف الموضوعى الهجائى الذى يعقب الجداول فى خطة التصنيف الوجهى المصطلح عليها. وتظهر المصطلحات مرتين – مرة فى المكنز ومرة فى الجداول ويربط بين المكانين رمز Notation – والمعلومات التى تعطى فى المكنز أكثر من المعلومات التى تعطى فى جدول التصنيف، ويزاد على هذه المعلومات ما يعطى فى جداول التصنيف، ولذا يعتبر الجزءان مكملين لبعضهما البعض – وإذا استعملا بمفردهما يعتبران غير كاملين. والمثال الآتى يوضح لنا كيفية الربط بين المصطلح والرمز. فلو بحثنا تحت Television Camera Tubes نجد الرمز MCE والذى يشير إلى مكان التصنيف فى الجداول، وبالرجوع إلى هذه الجداول نجد عرض Display للتركيبات الموضوعية التى ينتمى إليها اللفظ – ويتضح لنا أن العنصر الرئيسى للمصطلح هو Cathode ray tube والذى يتبع بدوره اللفظ الأوسع Electron Beam Deflection والأنابيب الإلكترونية Electron Tubes فى حين أن اللفظ الأضيق Narrow term (NT) هو Television Color Camera Tubes والتى تتضمن Television Picture Tubes, Image Converter Tubes .

فالتصنيف هنا يعرض هذه العلاقات التركيبية الموضوعية بطريقة أفضل من الطريقة المتبعة فى المكانز التقليدية خصوصاً فى المصطلحات الأوسع Broader Termes (BT) التى لا يكشف ترتيبها الهجائى عن المستويات الثلاثة المختلفة Broad term/Narrow Term/ Related Term/ BT/NT/RT.

٧- الكشافات المتسلسلة: Chain Indexes

يفشل المستخدمون للكشاف أحياناً فى العثور على مداخل مفيدة عند عدم تمكنهم من بحث أكثر الموضوعات المتخصصة أهمية .. ويحاول التكشيف المتسلسل التقليل من ذلك عن طريق تقديم المداخل المفردة فى الكشاف المصنف ــ واحداً بعد الآخر ــ فى قائمة هجائية .. ومعنى ذلك أن الكشافات المتسلسلة تضمن ربط كل مفهوم بالمفهوم المرتبط به مباشرة فى النظام الهرمى .. وذلك من العام للخاص، وحيث توجد جميع المصطلحات أو أسماء الموضوعات المكشفة.

هذا ويعتبر التكشيف المتسلسل واحداً من نتائج نظريات رانجاناثان للتصنيف، ومن الواضح أن التكشيف المتسلسل يساعد على التغلب على مشكلة المدخل العام Generic فى الكشاف المصنف، وذلك باستخدام جميع المصطلحات ذات الدلالة كنقاط مداخل، كما أن التكشيف المتسلسل يتكون بطريقة آلية وبالتالى يخفف العبء على المكشف من ناحية اتخاذ القرارات، وفى الأساس هو نظام تصنيف، كما يعكس وظائف النظم المصنفة خصوصاً من ناحية الاختيارات الإنسانية حيث يقرر القائم بالتجميع ما يجب أن يكون عليه النظام المنهجى للمعرفة الإنسانية.

ولتوضيح هذا التسلسل من العام للخاص يمكن التعرف على النتائج المتسلسلة لكتاب عن الاتحادات التجارية فى صناعة تعدين الفحم بيوركشير والمصنف حسب نظام ديوى العشرى برقم ٣٣١،٨٨١٢٢٣٣٠٩٤٢٧٤

٣٠٠ علوم اجتماعية

٣٣٠ اقتصاد

٣٣١ العمل والعمال

٣٣١،٨ * موضوعات أخرى

٣٣١،٨٨ اتحادات تجارية

٣٣١،٨٨١ * فى صناعات متخصصة

٣٣١،٨٨١٢ هندسة

٣٣١,٨٨١٢٢	تعدين
٣٣١,٨٨١٢٢٣	* أنواع التعدين
٣٣١,٨٨١٢٢٣٣	تعدين الفحم
٣٣١,٨٨١٢٢٣٣٠٩	* حسب المكان
٣٣١,٨٨١٢٢٣٣٠٩٤	أوروبا
٣٣١,٨٨١٢٢٣٣٠٩٤٢	بريطانيا
٣٣١,٨٨١٢٢٣٣٠٩٤٢٧	شمال وسط بريطانيا
٣٣١,٨٨١٢٢٣٣٠٩٤٢٧٤	يوركشير

وهنا يتم عمل مداخل الكشاف مبتدئين بالخطوة الأخيرة فى التقسيم مع استخدام نظام الببليوجرافيا البريطانية (B.N.B) للأسماء الجمع والتى تنفصل عن بعضها بالشارحة (:) أو الكولون كما يلى:

يوركشير: تعدين الفحم: الاتحادات التجارية.	٣٣١,٨٨١٢٢٣٣٠٩٤٢٧٤
بريطانيا: تعدين الفحم: الاتحادات التجارية	٣٣١,٨٨١٢٢٣٣٠٩٤٢
التعدين: الفحم: الاتحادات التجارية	٣١,٨٨١٢٢٣
التعدين: الاتحادات التجارية	٣١,٨٨١٢٢
الاتحادات التجارية	٣١,٨٨
العمل والعمال: اقتصاد	٣١
الاقتصاد	٣٠
العلوم الاجتماعية	٣٠٠

ويلاحظ أنه لم يتم عمل مداخل كشفية لما يلى (موضوعات أخرى ــ فى صناعات متخصصة ــ أنواع التعدين ــ حسب المكان) وذلك لأن أرقامها التصنيفية تعتبر مجرد رموز للربط أى لإدخال وجه أو قسم جديد للموضوع، كما

لم يتم عمل مداخل كشفية لما يلى (أوروبا ــ شمال وسط بريطانيا) نظراً لأن القارئ لن يستثمر الكشاف تحت هذه المصطلحات عادة، وإن اختلف علماء الفهرسة الموضوعية فى حفظ أو استبعاد هذه المصطلحات طبقاً لنوع المكتبة ومجموعاتها والمستفيدين منها إلخ. وواضح أن التكشيف المتسلسل يضم كلاً من التصنيف والتكشيف الموضوعى فضلاً عن القيمة الاقتصادية التى يوفرها التكشيف المتسلسل نظراً لأنه يستبعد المترادفات التى لا داعى لها.

وعلى ذلك فتكون القاعدة العامة هى:

(أ) استخدام المصطلح الأخير فى السلسلة ليكون الرأس المخصص والمباشر مع وصفه إذا دعت الحاجة بمصطلحات من الحلقات الأعلى فى السلسلة.

(ب) قم بعمل إحالة للرأس من الحلقة الأعلى مباشرة فى السلسلة وكذلك قم بعمل إحالة لتلك الحلقة من الحلقة الأعلى لها وهكذا ...

وإذا كان التكشيف المتسلسل قد استخدم فى بناء الكشاف الهجائى للببليوجرافيا البريطانية (B.N.B) من عام ١٩٥٠ وحتى عام ١٩٧٠، فقد بدأت المكتبة البريطانية عام ١٩٧٠ مجموعة من الإجراءات لتطوير النظام السابق وذلك بتطبيق نظام التكشيف المحافظ على السياق (بريسيس) Precis: The preserved Context indexing System .

كما سيأتى شرحه فيما بعد.

٨- نظام التحليل الكشفى للفهرس المصنف:

Indexing the Classified Catalogue

وهذا هو النظام المتبع فى مركز التوثيق التابع لمعهد التربية بجامعة لندن ويعتبر هذا النظام حصيلة بحث وعمل توثيقى كبير لأنه يجمع بين التحليل الكشفى والتصنيف فى نظام واحد، وقد اعتمد هذا النظام على الأفكار الأصيلة لرانجاناثان فى "تحليل الأوجه" والتى طبقتها العالمة باربارا كيل ... Barbara Kyle فى العلوم الاجتماعية على مدى سنوات طويلة ثم أفاد معهد التربية بجامعة لندن من هذا العمل فى العلوم الاجتماعية وطبقه فى علوم التربية وخصوصاً ما يتعلق منها بتنمية المجتمع.

وهذا النظام يعتمد على دوران رموز التصنيف Rotaion of Classification Symbols واستخدام رموز التصنيف ذاتها فى الترتيب بدلاً من الترتيب الهجائى Alphabetic وقد وضع هذا النظام كنظام يدوى شبيه بالنظم التقليدية المستخدمة بالمكتبات وكتطوير لها، كما أن هذا النظام يمكن استخدامه بواسطة الحاسب الآلى، ويعد له برنامج (شبيه ببرنامج الكشاف الدوار البرميوتى) حيث يتم فيه دوران المداخل الكشفية طبقاً لرموز التصنيف.

وهذ الفهرس يتشابه مع الفهرس البرميوتى فى الشكل ولكنه يختلف فى المحتوى ففى هذا الفهرس واسمه "الفهرس الدوار" Rotated or Cyclic فإن تجميع المداخل يتم طبقاً لنظام التصنيف دون الاعتماد على الكلمات المفردة .. فى العنوان أو فى النص .. وعلى سبيل المثال فإن موضوعاً مثل (تكوين مجموعات الكتب العلمية) يظهر عند الرمز الخاص "بالمعونة الفنية" رغم أن مصطلح "المعونة الفنية"لم يرد فى عنوان المقال ــ وذلك للدلالة على موضوع مرتبط ارتباطاً وثيقاً بهذا المقال..

٩- كشاف الخيط الرابط: String Indexes

يعتبر كشاف الخيوط الرابطة عادة ــ وليس دائماً ــ نتاج الحاسب الآلى، وفكرة هذا الكشاف هو عرض سلسلة من المداخل الكشفية الدوارة من قائمة أساسية من المصطلحات الكشفية التى تكون الخيط String .. والهدف من ذلك هو إعطاء المستفيد نقطة مدخل لجميع المصطلحات الكشفية وعرضها فى سياق بعضها البعض.

وعلى الرغم من أن تكشيف الخيط الرابط مصطلح حديث، إلا أن له سابقة محاولة الإفادة من الخطة التصنيفية والتى تعود إلى الأعمال النظرية لكل من فاراداان Farradane ورانجاناثان وكتر وغيرهم .. وقد عرف كرافن (Cravens, 1986) تكشيف الخيط الرابط **كما يلى**:

يعتبر كشاف الخيط الرابط شكلا من أشكال التكشيف له صفتان **أساسيتان وهما**:

١- كل مصطلح كشفى له عادة عدد من المداخل الكشفية تحتوى على الأقل على بعض المصطلحات نفسها.

٢- يولد برنامج الحاسب الآلى الجزء الوصفى من كل مدخل كشفى طبقاً لقواعد تركيبية Syntatical منتظمة واضحة، والجزء الوصفى لكشاف الخيط الرابط يسمى خيط الكشاف، برنامج الحاسب الآلى الذى ينتجه يسمى مولد خيط الكشاف String Index generator .

هذا ويمكن للمستفيدين تجنب البدايات الفاشلة منذ نقطة البداية، عن طريق ربط مصطلحات المداخل بغيرها من الواصفات الموضوعية، ذلك لأنه عن طريق ملاحظة المصطلحات المرتبطة يمكن معرفة أين يقف الباحث منها.

لنفترض أن لدينا موضوع عن "استخدام الحاسبات الآلية فى الرسوم المتحركة التليفزيونية فى الولايات المتحدة الأمريكية" .. فتحليل المفاهيم يعطينا خيطاً طوليًّا للمصطلحات الكشفية .. وهذه المصطلحات تغذى بالحاسب الآلى، وهذا ينتج بدوره مؤشرات للعلاقات المختلفة .. والحاسب الآلى يقوم بتدوير المصطلحات، آخذاً فى الاعتبار هذه العلاقات وبالتالى يقوم بإنتاج سلسلة من المداخل كما يلى:

- الولايات المتحدة
  صناعة التليفزيون. الحاسب، الرسوم المتحركة.
- صناعة التليفزيون، الولايات المتحدة.
  الحاسبات، الرسوم المتحركة.
- الحاسبات، صناعة التليفزيون، الولايات المتحدة، الرسوم المتحركة.
  الرسوم المتحركة، الحاسبات.

صناعة التلفزيون. الولايات المتحدة.

فتكشيف الخيط الرابط هو شكل من أشكال التكشيف الذى يستعين بالحاسب. الآلى، والقرارات الفكرية تتم بواسطة المكشف والعمل التطويعى يتم بواسطة الحاسب والبحوث جارية للتعرف على مدى إمكانية برمجة الحاسب الآلى ليتولى أجزاء أكثر وأكثر من العمل اليدوى.

هذا وتتراوح نظم تكشيف الخيط الرابط من مجرد تدوير الكلمات مع بعض القواعد، إلى النظم التركيبية المعقدة للغاية .. وفيما يلى بعض النماذج الأكثر تعقيداً

من نظم تكشيف الخيط الرابط وهو بوبسى Popsi ، نيفس NEPHIS ، سيفت CIFT وبريسيس PRECIS .

- بوبسى POPSI (Postulate - based Permuted Subject Indexing)

تطور هذا النظام فى الهند فى مركز بحوث التوثيق والتدريب .. وهو يتبع أفكار رانجاناثان فى التصنيف حيث يعتبر قواعد التكويد بسيطة نسبياً وشبيهة إلى حد كبير بتكشيف كووك KWOC مع بعض الإضافات.

نيفيس NEPHIS Nested Phrase Indexing Systems

وقد تم تطوير هذا النظام عل يد الباحث كرافن Craven حيث يصمم مدخل الخيط الرابط ليكون جملة Phrase فى اللغة العادية والتركيز فى نظام نيفيس على الاقتصاد فى الكلمات وعلى المكشف وعلى الباحث (Craven, 1986, 38).

سيفت CIFT

Contextual Indexing and Faceted Taxonomic Access System

وهو نظام تكشيف خيط رابط ذو غرض خاص، فقد طور لخدمة جمعية اللغات الحديثة (MLA)، حيث تنشأ مداخل موضوعية هجائية من الخيوط بواسطة المكشفين الذين يقومون بوضع وجوه Facets مستمدة من الأدب واللغة .. وقد نشر هذا النظام لاستخدامه مع البيليوجرافيا العالمية لجمعية اللغات الحديثة MLA)

البريسيس PRECIS Preserved Context Index System

لعل هذا النظام هو أفضل النظم المعروفة داخل إطار نظام تكشيف الخيط الرابط .. وهو مرتبط أيضاً بالحاسب الآلى وطبق فى البيليوجرافيا الوطنية البريطانية مع استخدام تسجيلات مارك وقد تطور هذا النظام لتقديم بيانات كشفية موضوعية لتسجيلات مارك، البريطانى وكذلك لإنتاج كشاف موضوعى هجائى للبيليوجرافيا الوطنية.

هذا ويعتبر نظام البرسيس PRECIS (Preserved Context Index) نقطة تحول جديدة فى مفهوم التحليل الموضوعى للمطبوعات. والغرض من وراء نظام البريسيس هو توفير كشاف هجائى موضوعى لتسهيل مهمة البحث فى أشرطة

المارك MARC عن طريق استخدام أى مصطلح له وزنه فى الكشاف ثم التدرج من هذا المصطلح إلى المضمون الذى ورد فيه هذا المصطلح. ويختلف نظام البريسيس عن نظم التكشيف الآلية المعروفة فى أنه يستخدم سطرين Two- Line Format على عكس نظام السطر الواحد one- Line الذى يستخدم فى رؤوس الموضوعات الموجودة فى الكشاف المتسلسل Chain index والكلمات الدالة فى السياق (كويك) KWIC .

والمثال التالى قد يشرح لنا كيفية تطبيق نظام البريسيس

فمدخل البريسيس PRECIS لوثيقة بعنوان A guide for staffing a

Hospital nursing Service Patients

Hospital Patients. Nursing. Personnel

هذا المدخل ذو السطرين مبنى على المفهوم المعروف بتطويع الخيط Manipulation String حيث تحدد أدوار المصطلحات وعلاقتها بعضها ببعض حتى تحدد العلاقة اللغوية بين كل مصطلح وآخر.

مثال من تطويع الخط : **Manipulation String**

$21101 $ a patients $ c Hospital $ Z101 $ a nursing $ Z022 $ a hospital nursing $ 101 $ a personnel

Labour .... Personnel

Staff .... Personnel

هذا وكل واحد من هذه المصطلحات يحتل مركز بداية Lead ويدفع المصطلح الآخر لمركز البداية وذلك كما يلى:

Patients

Hospital Patients. Nursing. Personnel

Hospital Patients

Nursing. Personnel

Nursing. Hospital Patients

Personnel

Personnel. Hospital Nursing
Patients

هذا ويجب الوصول إلى اتفاق على أحسن الطرق لعرض الخيط String فى البطاقة ولأسباب اقتصادية وشكلية يفضل أن يطبع الخيط مع كافة الرموز فى الحاشية تاركين لكل مكتبة حرية تفسير هذه البيانات لطبعها فى المداخل الموضوعية على البطاقات. مثل هذا الخيط مع البيانات الرمزية (الكودية) الكاملة ستكون بهذا الشكل:

$ z11030 $ a digital Computer systems $ Z 21010 $ a programming $ z 31030 $ a Basic Language

وهذا المثال يوضح لنا أن على المفهرسين وغير المهنيين أن يتعلموا نظام البريسيس حتى يتمكنوا من شرح هذه البيانات بدقة على النحو الآتى:

Digital computer systems

1 Programming. Basic Language

Programming. Digital Computer Systems
Basic Language

Basic Language. Digital Computer Systems
Programming

كما أن هذا المثال يوضح لنا حتمية إنشاء فهرس موضوعى ثانى إلى جانب الفهرس الذى تستخدمه المكتبة (مثل فهرس مكتبة الكونجرس) وفى الفهرس الثانى توضح بطاقات البريسيس نظراً لاختلافها شكلاً ومفهوماً عن أى فهرس موضوعى آخر.

وأخيراً فتقدر تكاليف إدخال البريسيس بحوالى مليون دولار سنوياً وذلك بسبب ضرورة زيادة عدد المكشفين والمحررين.

هذا ومداخل البريسيس Precis وأرقام تصنيف ديوى العشرى موجودة حالياً على أشرطة مارك الإنجليزية وبالتالى يسهل على المكتبات الأخرى استخدام هذا النظام التكشيفى. كما أنه نظراً لأن البريسيس Precis يستخدم خيطاً واحداً للمصطلحات Single String Terms لخلق مداخل متعددة .. فإن استعماله سيؤدى إلى تسهيل عملية البحث الموضوعى فى المارك Marc .

١٠- كشافات الاستشهادات المرجعية: Citation Indexing

يعود مفهوم كشافات الاستشهادات إلى أكثر من قرن من الزمان حيث استخدمها المحامون فى بريطانيا فيما يسمى باستشهادات شبرد Shepard والتى تعطى نبذة عن القوانين الصادرة ثم الاستشهادات التى تمت بها فيما بعد.

وواضح أن هذه أداة بحثية أساسية بمهنة المحاماة نظراً لاعتمادها على السوابق Precedent ومع ذلك فتكشيف الاستشهادات أداة مرجعية عامة جديدة نسبياً. وكما رأينا فمعظم الكشافات تتبع إجراءات اختيار المصطلحات الكشفية من الوثائق أو أنها تختار مصطلحات المفاهيم من قائمة مصطلحات ثم ترتب المصطلحات لإظهار المحتوى الموضوعى للوثيقة.

ولتحسين هذه الكشافات استخدمت أدوات دلالية (معنى) وأدوات تركيبية (اللغة) وذلك لتوضيح العلاقات فى الطريقة التى تستخدم بها المصطلحات.

أما تكشيف الاستشهادات فهو يتبع مدخلا مختلفاً تماماً، فهو لا يعتمد على كلمات كشفية، وبالتالى يتجنب المشكلات الفكرية الخاصة بالمعانى وتفسيرها فى التكشيف التقليدى.

والافتراض الأساسى فى تكشيف الاستشهادات، أننا نعتمد على مؤلف الوثيقة ليدلنا على الموضوع أى أننا لا نعتمد على المكشف، ومعنى ذلك أننا فى كشافات الاستشهادات نرى المؤلف أكثر الأشخاص تأهيلاً فى تحديد وتفصيل المواد المتعلقة بموضوعه، هذا وقد صدر عن طريق معهد المعلومات العلمية فى أمريكا المجموعات الثلاثة التالية:

- Science Citation Index (SCI) ١٩٦١
- Social Sciences Citation Index (SSCI)١٩٦٩

Arts ands Humanities Citation Index (AandHCI) ١٩٧٨ -

والفكرة وراء إصدار هذه الكشافات هو أن هناك علاقة موضوعية بين الأعمال العلمية السابقة والأعمال التى استشهدت بها بعد ذلك، ويتكون كل واحد منها من ثلاثة كشافات فرعية وهى:

١- كشاف الاستشهادات Citation Index

٢- كشاف الوثائق المصدرية Source Index

٣- كشاف التباديل الموضوعى Permutation Subject Index

ويرتب الكشاف الفرعى الأول حسب أسماء مؤلفى الوثائق المستشهد بها وذلك فى ترتيب هجائى حسب المؤلفين.

وفى حالة تعدد الوثائق التى تستشهد بنفس المؤلف الأول ترتب هذه الوثائق زمنياً وفقاً لتواريخ نشرها وتستكمل البيانات الببليوجرافية.

أما كشاف الوثائق المصدرية فهو مرتب هجائياً حسب أسماء مؤلفى الوثائق التى وردت بها الاستشهادات وبعدها البيانات الببليوجرافية الكاملة.

أما كشاف التباديل الموضوعى فيتم بواسطة الحاسب الآلى وذلك بإعادة ترتيب الكلمات الدالة الواردة فى عناوين الوثائق المصدرية وذلك وفقاً لمختلف الأوجه الممكنة للمصطلحات، وليس مجرد ترتيب كلمات العنوان بالتناوب كما هو الحال فى كشاف الكويك السابق الإشارة إليه.

هذا ويحتفظ معهد المعلومات العلمية فى فيلادلفيا بالتسجيلات الإلكترونية لهذه الكشافات الثلاثة ويمكن استخدامها على الخط المباشر أيضاً وهناك شك من قبل بعض الباحثين فى أن تكون هذه الكشافات مقياساً لكفاءة الباحثين وقدراتهم الإنتاجية فى مجال بحثهم .. ولكن هذه المراصد الإلكترونية لكشافات الاستشهادات المرجعية تعتبر مصدراً هاماً فى دراسة الخصائص البنائية وطريقة النمو للإنتاج الفكرى وكان لتحليل هذه الاستشهادات أثره فى نمو دراسات القياسات المعلوماتية.

ثالثاً : خطوات عملية التكشيف وخصائصها:

(أ) الهدف من الكشاف ومشكلاته :

الهدف من الكشاف هو استرجاع الوثائق أو التسجيلات التى اختزنت ونظمت بواسطة عملية تحليل المحتوى الموضوعى لأوعية المعلومات، والتعبير عن هذا المحتوى- بلغة نظام التكشيف، والمقصود بتحليل المحتوى هنا هو إعداد المداخل والروابط التى توصلنا للمعلومات فى مصادرها. والمدخل يمكن أن يكون كلمة أو جملة مفتاحية أو رمز وذلك للدلالة على مكان المادة أو المواد المرتبطة بمفهوم معين، إن عدم التكشيف الجيد ينتج عنه استرجاع وثائق غير مفيدة كما قد ينتج عنه تجاهل وثائق مفيدة للباحث. ولما كان استرجاع المعلومات يعتمد على المقارنة أو المضاهاة المنطقية لمصطلحات كل من السؤال ومصطلحات تكشيف الوثائق فمن اللازم الإشارة إلى ما يلى:

- يحتاج الباحث إلى التعرف على المبادئ المستخدمة فى التكشيف.
- يجب أن يضع المكشف فى اعتباره المستخدمين المحتملين للكشاف.

هذا ويعتمد الكشاف عادة على سلسلة من رءوس الموضوعات أو المداخل أو الواصفات المرتبة فى نظام معين .. وكل واحدة من المداخل مرتبطة بالمداخل الأخرى ذات العلاقة .. فعلى سبيل المثال فإن كشاف المؤلفين يحتوى على أسماء المؤلفين مرتبة هجائية، ويصحب كل واحد من هذه الأسماء رقم الوثيقة كمرجع أو أرقام التصنيف واستشهادات الوثائق، لكن اهتمامنا الرئيسى هو بالكشافات الموضوعية وتلك التى تعتمد على رءوس الموضوعات الهجائية.

وهناك بعض الكشافات المطبوعة والمنشورة مثل الكشاف الهندسى Engineering Index وهناك كشافات أخرى يتم إعدادها داخل مراكز المعلومات لاستخدام العاملين بالمركز، وهناك كشافات كذلك تختزن فى الحاسبات والحاسبات المصغرة، وفى هذه الحالة الأخيرة فإن البحث يجب أن يتم عن طريق الاتصال بذاكرة الحاسب عبر النهايات الطرفية Terminal هذا وتوفر لنا نظم المعلومات المعتمدة على الحاسبات أشكالاً واسعة للتكشيف، قد لا يمكن توفيرها بالتكشيف اليدوى أو المطبوع، وكل منهما ـــ أى المحسب واليدوى ـــ له عيوبه ومزاياه

وسنتناولها فيما بعد، وما يهمنا فى هذا الاستعراض أن نبين أن جميع الكشافات الجيدة تشترك فى وجوه معينة، كما أن الكشاف ــ مهما اختلفت النظريات التى وراءه ــ يجب أن يكون مفهوماً لأغلبية المستفيدين.

وهناك نوعان من المشكلات التى تنبثق من هذه الحاجة الأساسية للمستفيدين وهما:

- يجب على المكشف تحديد الموضوع بكفاءة وفعالية.
- يجب على المكشف أن يجد الطريق للدلالة على العلاقات بين الموضوعات حتى يمكن القيام باستراتيجية البحث.

(ب) منهجية عملية التكشيف:

تتضمن عملية التكشيف ثلاثة خطوات مبدئية على الأقل وهى: القراءة الواعية >التحليل> تحويل المفاهيم إلى مصطلحات كشفية والهدف من القيام بهذه الخطوات هو بناء سمات الوثيقة Document Profile (والسمات هنا هى مجموعة كاملة من مفاتيح البحث الممكنة)، أما بالنسبة للقراءة الواعية فالهدف منها إحاطة المكشف بالمحتوى الموضوعى للوثيقة بنفس الطريقة التى يغمس بها القائم بالاستخلاص نفسه فى الوثيقة. كما ينبغى على المكشف أن يتعرف على إطار وتركيب الموضوع الذى يقوم بتكشيفه وما يمكن لهذه الوثيقة أن تصنعه بالنسبة لتقدم المعرفة. ومن الواضح أن هذه الخطوة لا تتم فى نظم التكشيف المعتمدة على الحاسب الآلى حيث هناك مصطلحات تحدد بواسطة الآلة Machine assigned terms .

أما بالنسبة للتحليل فهى أصعب الخطوات لأنه مرحلة علمية وفنية فى ذات الوقت، فبعض جوانب عملية التحليل يمكن تحديده ولكن بعضها الآخر يعتمد على الخبرة والفطنة Intuition ، أى أن النظرية التى تقوم بدراسة عملية التحليل تنبع من الخبرة وليس العكس .. فالاتجاه العام للوثيقة يجب أن يكشف ولكن إلى أى حد سيقوم المكشف بتكشيف الاتجاهات الثانوية أو ذات الأهمية الأقل؟ لعل ذلك يتحدد بالاستخدام الفعلى للكشاف النهائى (على صعوبة توقع ذلك). وعلى كل حال فهناك بعض القواعد المرشدة فى كل مطبوع كشفى أو استخلاص يستعين به المكشف أثناء

عمله. وعند الانتهاء من التحليل المبدئى للمفاهيم الموجودة فى الوثيقة، فإن المكشف يقوم باختيار المصطلحات التى تضاهى المفاهيم التى يجب تكشيفها .. وما ينبغى الإشارة إليه هنا أن هذه الخطوات الثلاث يمكن أن تتم فى نفس الوقت وليس بالضرورة بالتتابع كما سبق بيانه.

(ج) **خطوات التكشيف** اقترح كليفيلند (Cleveland, D., 1990, 104): **الخطوات التالية للتكشيف المنطقى:**

١- **تسجيل البيانات البيليوجرافية:** بدقة وذلك لأن التسجيل غير الدقيق سيجعل الوثيقة نفسها غير متاحة.

٢- **تحليل المحتوى:** ليس هناك ضرورة لقراءة كل وثيقة بالكامل، فالقراءة الاستطلاعية Spot Reading قد تكون كافية للمكشف للتعرف على المفاهيم الأساسية وذلك فى نطاق تحقيق السياسة التحريرية المتفق عليها، فقد تكون التوجيهات بالنسبة للإنتاج الفكرى العلمى التركيز على المنهجية والقياس والتجهيزات والنتائج الجديدة مع إهمال المواد التاريخية .. كما قد تشمل التوجيهات استخدام كلمات من النص أو من مكنز معين أو منهما معاً. هذا وتأتى المفاهيم المطلوبة من العنوان خصوصاً تلك المتعلقة بالموضوعات الرئيسية، أو من المستخلص كبديل للوثائق خصوصاً عند معالجتها للموضوعات المفتاحية بالوثيقة ... أو من النص نفسه حيث تقسم أوراق البحوث عادة إلى مقدمة وملخص ونتائج فضلاً عن العناوين الفرعية ..ويجب الاهتمام بالفقرات الأولى والأخيرة من المقال، وأخيراً فالقسم الخاص بالمراجع يسهم فى التعرف على تحليل المحتوى الموضوعى عن طريق تكشيف الاستشهادات والتزاوج الببليوجرافى.

٣- **تحديد الموضوع:** أى التعبير عن المفاهيم بالوثيقة بواسطة قائمة واصفات Descriptors ولكن اختيار الموضوعات الهامة يرتبط بشدة بسياسات التكشيف التى تضعها هيئة معينة، فضلاً عن أن المادة الموضوعية لكل وثيقة يتم تقييمها بناء على المتطلبات المعلوماتية للمستفيدين وهم الذين يحددون الموضوعات الهامة بالوثيقة، وهذا ما يسهل على المكشفين كإطار فكرى.

٤- **تحويل قائمة المفاهيم إلى لغة نظام التكشيف:** إذا كانت المصطلحات هى المستخدمة، فإن المكشف سيترجم المفاهيم إلى مصطلحات كشفية معيارية محكومة، يتم مضاهاات على المكنز أو أى قائمة استناد أخرى.

٥- **تقييم وفحص:** ما سبق القيام به من خطوات للتأكد من أن الواصفات التى وضعت تغطى جميع المفاهيم الهامة التى تعبر عن الموضوع المطلوب بدقة.

وتتميز عملية التكشيف بعدة خصائص أهمها الشمول والتخصيص والمقصود بالشمول Exhaustivity إبراز الموضوعات أو الأفكار الموجودة بالوثيقة، وهذا يؤدى إلى زيادة عدد المداخل المعبرة عن التحليل نسبياً .. وقد استخدم الكاتب عبارة الزيادة النسبية للمداخل نظراً لأن عملية التكشيف انتقائية تعتمد على اهتمامات المستفيدين من النظام، وطبيعة الوثائق نفسها، وطريقة إخراجها، فضلاً عن طبيعة المكشف أى صفاته الشخصية كالحرص والدقة والصبر، ومن حيث اهتماماته الموضوعية.

ومع ذلك فنظم التكشيف تضع بعض المعايير ومقننات للالتزام بها، كأن يوضح المكشف المصطلحات المختارة والتسلسل المنطقى للمشكلة الرئيسية (المصطلحات المحورية أو البؤرية) المشاكل الفرعية (المصطلحات الفرعية ثم الهامشية) ثم منهج الدراسة وطبيعتها ثم النتائج الأساسية والفرعية.

كما أنه من المعروف أنه كلما ازداد مستوى الشمول ازداد عدد الوثائق المسترجعة (Recall) استجابة لسؤال معين، ولكن هناك تناسب عكسى بين الاستدعاء Recall والدقة Precision وذلك طبقاً للمقياس الرياضى التالى الذين يحدد فاعلية النظام.

$$\text{مقياس الاستدعاء} = \frac{\text{أ}}{\text{أ} + \text{ب}} \qquad \text{مقياس الدقة} = \frac{\text{أ}}{\text{أ} + \text{جـ}}$$

$$\text{مقياس فاعلية النظام} = \frac{\text{أ}}{\text{أ} + \text{جـ}} + \frac{\text{د}}{\text{د} + \text{جـ}}$$

حيث أن (أ) هى الوثائق المفيدة المسترجعية، (ب) هى الوثائق المفيدة غير المسترجعية و(جـ) هى الوثائق غير المفيدة المسترجعية، (د) هى الوثائق غير المفيدة وغير المسترجعية.

وهذا كله عن مشكلة الشمول، أما خاصية التخصيص Specificity فتدل على مدى تطابق المصطلحات الكشفية مع موضوعات الوثيقة ووصفها بطريقة محددة، وإذا لم يتوافر شرط التخصيص هذا أصبح المصطلح الكشفى الواحد يستعمل فى تكشيف عدد كبير من الوثائق المختلفة عن بعضها البعض.

وهناك ارتباط بين التخصيص فى التكشيف ومعدلات الاستدعاء والدقة، ذلك لأنه كلما كانت لغة التكشيف أكثر تحديداً، ارتفع معدل الدقة لأن الوثائق المطابقة للمصطلحات الكشفية المخصصة هى التى ستسترجع فقط.

الفصل الثالث

اللغة الطبيعية والمصطلحات المحكومة في استرجاع المعلومات [1]

تقديم :

يحفل الإنتاج الفكرى للمعلومات والمكتبات بمقارنات عديدة بالنسبة لمزايا وعيوب كل من اللغة الطبيعية والمصطلحات المحكومة فى تشغيل

نظم استرجاع المعلومات، ويذهب البعض إلى أن اللغة الطبيعية تعتبر ذات أهمية بالنسبة للموضوعات الجديدة وبالتالى ستحسن كلاً من الاستدعاء والدقة Recall and Precison أى أنها اللغة المفضلة للمتخصص الموضوعى، ويذهب البعض الآخر إلى أنه على الرغم من الإمكانات المحتملة الكبيرة لاستخدام اللغة الطبيعية فى الحاسب الآلى بواسطة المستفيدين، إلا أننا فى المرحلة الراهنة على الأقل يجب أن نركز على المصطلحات المحكومة التى تعالج المترادفات وتعالج مشكلات الجناس homography وتربط بين المصطلحات المتقاربة، وبالتالى تقلل من عملية تخمين الكلمة الصالحة لدى المستفيد، أى أن اللغة المحكومة ستفرض عملية الانتظام أى تمثيل المحتوى الموضوعى بين الوثائق وتمدنا بمصطلحات المفاهيم العريضة التى لا تكون موجودة فى النص، وعن طريق التركيب الهرمى والإحالات تقدم للمستفيد معاونة إيجابية فى تحديد المصطلحات البحثية المناسبة.

وهناك فريق ثالث يحاول التوفيق بين وجهتىْ النظر السابقتين، وذلك باستخدام اللغة المهجنة hybrid Language نظراً لأن بحوث النص الحر أو اللغة الطبيعية يمكن أن تسترجع بعض المواد الصالحة relevant التى لا يمكن الحصول عليها ببحوث اللغة المحكومة والعكس صحيح، وبناء على ذلك فقد ذهب بعض المؤلفين إلى إمكان تحسين بحث اللغة الطبيعية عن طريق مكنز البحث أو مصطلحات التحكم اللاحق Post-Controlled vocabulary ، حيث يرى هذا الفريق أن مثل

---

(١) نشر هذا البحث قبل التعديل الحالى فى: مجلة مكتبة الملك فهد الوطنية. مج٣، ع١ (اكتوبر ١٩٩٧).

هذا النظام سيقدم لنا جميع ميزات اللغة الطبيعية مع بعض مميزات مصطلحات التحكم القبلى.

وستتناول الدراسة التى بين أيدينا المحاور التالية:

أولاً : بعض التعاريف.

ثانياً: التطور التاريخى لاستخدام اللغة الطبيعية فى الاسترجاع اليدوى والمحسَّب.

ثالثاً: الحاجة إلى المصطلحات المحكومة.

رابعاً : مقارنة بين اللغة الطبيعية ولغة المصطلحات المحكومة فى استرجاع المعلومات.

خامساً: البحوث المستقبلة فى كل من المعالجة الآلية للغة الطبيعية وبناء المكانز الآلية.

أولاً : بعض التعاريف :

١- لغة التكشيف

لا تعتبر لغة التكشيف مجرد قائمة من مصطلحات التكشيف المقبولة لدى المستفيدين، ذلك لأن لغة التكشيف تحتوى على آلية لبناء واستخدام هذه المصطلحات، أى أنها تشكل قواعد استخدام المصطلحات فى علاقة المصطلحات ببعضها، هذا وتخدم بعض المصطلحات فى إرشاد المستفيد للمصطلحات المضبوطة التى يمكن أن يستخدمها المكشف.

وهذه المصطلحات المؤشرة Pointer إلى جانب مصطلحات التكشيف المسموح بها تشكل ما يمكن تسميته **بمصطلحات المداخل** Entry Vocabulary أى أن مصطلحات المداخل الكلية تحاول التنبؤ بدقة ــ على قدر المستطاع ــ بجميع الطرق الممكنة التى توصل المستفيد بقاعدة المعرفة، ومصطلحات المداخل هذه تساعد كلا من المستفيد والمكشف على اختيار المصطلح الصحيح، وتقلل بالتالى من الجهد الفكرى اللازم لاتخاذ القرار.

ويمكن تقسيم لغات التكشيف إلى نظم مصطلحات معينة assigned- term systems ونظم مصطلحات مشتقة derived - term systerm ففى النظم الأولى يقوم المكشف بتعيين أو وضع مصطلحات أو واصفات على أساس التفسير الذاتى للمفاهيم الواردة فى الوثيقة وهو حين يفعل ذلك فهو يبذل بعض الجهد الفكرى، ذلك لأن المكشف يحدد موضوع الوثيقة كما يراه ثم يقرر ما هى المصطلحات المناسبة التى يتم ترشيحها. أما فى النظم المشتقة فجميع الواصفات تؤخذ من المادة نفسها.

وبالتالى فإن كشافات المؤلفين والعناوين وكشافات الاستشهادات المرجعية وكشافات اللغة الطبيعية تعدّ نظماً مشتقة، بينما تعد جميع لغات التكشيف ذات أدوات ضبط المصطلحات كقوائم رءوس الموضوعات والمكانز وخطط التصنيف، هذه جميعاً تعد نظم مصطلحات معينة assigned ، وهذه تتضمن جهداً فكرياً أكثر من غيرها، وبالتالى فهى مكلفة وتستغرق وقتاً أطول، أما النظم المشتقة فهذه يتم التفكير فيها على اعتبار إمكان القيام بها آلياً مثل كشافات كويك Kwic حيث الاختيار بواسطة الحاسب الآلى، وقد تسمى التكشيف بالاقتباس Indexing by extraction .

وطبقاً لما يذهب إليه المدافعون عن النظم المعينة، فإن ضبط المصطلحات سيؤدى إلى انتظام التكشيف وكفاءته، ولكن الاختبارات العديدة التى أجريت على نظم ولغات التكشيف خلال السنوات الماضية لا تؤيد هذا الدفاع. (cleveland, D. 1990, 78)

٢- اللغة الطبيعية وبحث النص الكامل :

يمكن اعتبار اللغة الطبيعية مرادفة للغة المستخدمة فى الكتابة والحديث واللغة المحكومة عكسها، وفى إطار استرجاع المعلومات فإن مصطلح اللغة الطبيعية يدل عادة على الكلمات الموجودة فى النص المطبوع، وبالتالى فإن النص الحر Free Text يمكن اعتباره كأحد مرادفات اللغة الطبيعية، والنص الحر يمكـــن أن يتكـــون من:

أ – العنوان ب- المستخلص ج – الاقتباس د- النص الكامل للوثيقة

ونحن نعنى بإمكان بحث النص الكامل – إمكان اختزان النص الكامل فى شكل مقروء آلياً، ثم البحث عن كلمات أو توليفات أو ارتباطات للكلمات بواسطة الحاسب الآلى، بينما يمكن اعتبار النص الحر كجزء متماسك من النص الأصلى أو بعض الكلمات أو الجمل المقتبسة من النص بواسطة المكشف الإنسان (أو بواسطة برنامج الحاسب الآلى)، ثم إضافاتها للتسجيلة الببليوجرافية بوصفها تمثل النص. وفى بعض الأحيان فإن هذه المصطلحات المقتبسة تضاف لعناوين المواد المكشفة وتشكل عناوين أكثر ثراءً أو أكثر إيضاحاً (Lancaster, 1991, 193).

٣- مصطلحات ونظم التصنيف واسترجاع المعلومات:

تستخدم مصطلحات التصنيف عادة فى نظم المكتبة التقليدية مثل التصنيف العشرى العالمى وتصنيف مكتبة الكونجرس .. واستخدام هذه النظم محدود لأن هذه التصانيف عريضة فى نطاقها وتغطى مساحات واسعة من المعرفة، وبالتالى فهى لا تشمل المجالات المتخصصة المحددة التى تعالجها لغات التكشيف، أى أنها تفتقر إلى التعاريف الدقيقة فضلاً عن افتقارها للمرونة.

لقد اعتقد المكشوف فى وقت من الأوقات أن هناك تمييزاً واضحاً بين المكنز والتصنيف التقليدى، حيث كان المكنز يستخدم فى نظم الربط الترابط السابق ... ومع ذلك فقد تطور المكنز مع السنين ليكون أداة أكثر تعقيداً، تظهر لنا العلاقات الهرمية فضلاً عن ميزات وصفات التصنيف الوجهى، والآن يستخدم المكنز فى نظم الترابط السابق بينما تؤدى خطط التصنيف دوراً فى نظم الترابط اللاحق (Cleveland, D. 1990, 85)

لقد اعتمدت نظم التصنيف فى الماضى على السند الأدبى أو العلمى أو الفلسفى أو التعليمى ... وفى الوقت الحاضر هناك اهتمام دولى بتأثير التكنولوجيا الالكترونية على التصنيف وعلى الفهرسة الموضوعية، وتأثير ذلك على دول العالم الثالث، كما أن هناك مشروعات عديدة تحاول التعرف على قيمة خطط التصنيف فى الاسترجاع خصوصاً بالنسبة للفهرس العام المباشر OPAC (Online Public Access Catalog) ويتناول بعض الباحثين كلاً من نظام الكونجرس (LCC) ونظام ديوى العشرى (DDC) ومدى فائدتهما لهذا الغرض

مستخدمين فى ذلك فهرس الخط المباشر مارك OCLC MARC أى مارك الموجود فى فهارس الخط المباشر فى مركز المكتبة المحسب فى أوهايو كقاعدة للبحث، وقد أظهر ماركى (Markey, K. 1987) أن استخدام نظام ديوى العشرى للاسترجاع يمكن أن يزيد إمكاناته فى الاسترجاع إذا ما استخدم مع رءوس موضوعات مكتبة الكونجرس (LCSH) وكذلك الكلمات المفتاحية فى العنوان ... وهناك من يبحث فى العلاقة بين طول أرقام التصنيف (فى كل من تصنيف مكتبة الكونجرس وتصنيف ديوى العشرى) وخصوصية الموضوع Subject Specificity أما جوبينات (Gopinath, 1987) فقد تناول استخدام خطط التصنيف والمكانز فى علاقة تكافلية Symbiotic Relationship من شأنها تحسين استرجاع المعلومات .. ولعل مفهوم المكنز الوجهى Thesaurofacet أو Classaurus يعد أداة تضم خصائص خطة التصنيف الوجهية الهرمية مع خصائص المكنز، وهى معبرة عن التطبيق العملى لهذه العلاقة التكافلية.

وهناك من الباحثين من يتناول اثنين من المكانز التى استخدمت تصنيف بليس كمصدر لإنشاء المكانز خصوصاً باستخدام الطبعة الثانية من تصنيف بليس (Lancaster, F, 1989) .

٤- رءوس الموضوعات واسترجاع المعلومات:

لقد كان لظهور الفهرس العام المباشر (OPAC) أثره فى استعادة الاهتمام برءوس الموضوعات، كما تستخدم تقليدياً فى المكتبات فضلاً عن التعرف على الصعوبات فى استخدام المصطلحات اللازمة للإتاحة الموضوعية الفعالة فى الوسط الإلكترونى، ويعد ماركى (Markey, 1988) من بين الدارسين لأهمية دمج قائمة رءوس موضوعات مكتبة الكونجرس (LCSH) فى الشكل الإلكترونى ضمن الفهرس العام المباشر.

وعلى مر السنين تجمع نقد كثير لقائمة رءوس موضوعات مكتبة الكونجرس (Cochrane, P 1986) ومن بين هذا النقد ما يلى:

أ – الحاجة إلى تحسين شكل رأس الموضوع وإضافة التبصرات (Scope notes) أكثر.

ب- الحاجة إلى تحسين هيكل الإحالات التبادلية Cross-references.

ج - الحاجة إلى تحسين درجة الوصول فى الأقسام الفرعية.

د - الحاجة إلى تحسين البناء الرابط syndetic للنظام باستخدام التصنيف، هذا فضلاً عن أن قائمة رءوس موضوعات الكونجرس قد اتخذت منذ البداية موقفها بالنسبة للاعتماد على السند الأدبى Literary Warrant وليس على الإفادة من البحوث العلمية ومصطلحاتها المتطورة.

وحتى عندما قامت مكتبة الكونجرس بتحويل القائمة إلى مكنز، فقد كان هناك نقد بالنسبة للمصطلح "الأوسع" والمصطلح الأضيق و"المصطلح القريب" related term وذلك لأن القائمة لم تطبق العلاقات المنطقية الدقيقة التى تعنيها هذه المصطلحات، وبالتالى يمكن عدّها مجرد قائمة برءوس الموضوعات المستخدمة فى مكتبة الكونجرس وليس مكنزاً حقيقيّاً.

٥- ضبط المصطلحات وآليات المكنز :

الهدف الرئيسى من ضبط المصطلحات هو تعزيز الاتفاق المفهومى بين كل من المكشف والمستفيد من الكشاف، فالمكشف يفحص الوثيقة عقليّاً وفكريّاً للتعرف على ما يقصده المؤلف فى كتاباته ثم يقوم المكشف باختيار بعض المصطلحات من المكنز بحيث تمثل المصطلحات المختارة المفاهيم والعلاقات المناسبة للنص الأصلى ... أما المستفيد فيقترب من الكشاف بمفاهيمه ومصطلحاته الشخصية، وبالتالى فوظيفة آليات المكنز هى فى الواقع قيادة وإرشاد كل من المكشف والمستفيد إلى نقطة البحث نفسها... فضبط المصطلحات نشاط معقد، وهو واحد من القضايا الأساسية فى التكشيف (Cleveland, D., 1990, 77) .

ثانياً: التطور التاريخى لاستخدام اللغة الطبيعية فى الاسترجاع اليدوى والمحسب:

لقد كان للاستخدام الأولى للغة الطبيعية أو المصطلحات الحرة فى كشافات الترابط ومشتقاتها .. وهذه المصطلحات يتم اختيارها بواسطة المكشفين حيث يعتمدون على خلفياتهم العلمية فى التعرف على معانى الكلمات بالنص، وقد يستعينون فى ذلك ببعض القواميس المعيارية أو المكانز العامة لتحديد التعاريف ..

وكلما زاد عدد الوثائق التى يتم تكشيفها اتسعت تعاريف المكشف وأصبحت أكثر تخصيصاً .. وكلما نمت مصطلحات الكلمات الحرة وكلما تم تفسير مجموعة الوثائق بطرق مختلفة – دون أية محاولة لضبط المصطلحات – فإن ذلك ربما يؤدى إلى فوضى مما يجعل الكشاف نفسه لا فائدة منه، وقد يؤدى ذلك إلى نوع من تكشيف اللغة الطبيعية المرتبط بطريقة غير مباشرة بالمكنز خصوصاً من جانب المستفيدين الذين قد يستعينون بمكنز موضوعى مقبول للعثور على المصطلحات المطلوبة للوصول إلى المعلومات التى يحتاجونها فى مرحلة البحث والسيطرة على مشكلات المترادفات والعلاقات العامة Generic relationship وبالتالى تجميع المصطلحات القريبة مع بعضها.

وفى الوقت نفسه هناك جهود مستمرة لتحسين مداخل اللغة الطبيعية واستخدامها فى قواعد المعلومات والنهايات الطرفية على الخط المباشر مع تطور موازٍ للإفادة من التطورات التكنولوجية وجعل الاختزان والبحث فى النص الكامل أمراً ممكناً (Clevelend, D, 1990, 79) أى أننا نستطيع أن نرد الاتجاهات الحديثة فى استخدام اللغة الطبيعية فى استرجاع المعلومات إلى نظام المصطلح الموحد Uniterm الذى ابتدعه مورتيمر تاوبيه M. Taube عام ١٩٥١م؛ وقد كان للقواعد التى وضعت لنظام المصطلح الموحد جاذبية فى وقتها، ذلك لأن الجانب الموضوعى للوثائق يمكن تمثيله بدرجة كافية بواسطة الكلمات الواحدة أو الموحدة المقتبسة من نص الوثائق بواسطة المكشفين العاديين، ويُستكمل النظام عن طريق وجود بطاقة لكل مصطلح، ويوضع تحت هذا المصطلح أرقام الوثائق بطريقة معينة، بحيث يمكن إجراء البحوث عن طريق مقارنة الأرقام فى بطاقتين أو أكثر.

لقد كان للعالم Taube تأثيره الواضح على تطور نظم استرجاع المعلومات فى الخمسينيات، ولسوء الحظ فإن نظام المصطلح الواحد كان أقل جاذبية مما ظهر عليه الحال لأول مرة، ذلك لأنه واجه مختلف مشكلات المصطلحات المضبوطة أو المحكومة التى كانت تنتظر الحل، فقد تبين أن الموضوعات المرتبطة ببعضها ظهرت تحت كلمات واحدة، مختلفة، وأن البحث الشامل عن موضوع معين سيتطلب من الباحثين التفكير فى جميع الطرق التى يمكن بها تمثيل هذا الموضوع فى النص، وهذا أمر ليس باليسير من غير شك، وقد أدت

هذه المشكلات إلى الرجوع إلى المصطلحات المحكومة، لتطوير مكانز استرجاع المعلومات، وقد واجه نظام المصطلح الواحد مشكلات عديدة إلى جانب مشكلات المصطلحات السابقة، فالباحث الفرد مثلاً يمكن أن يقارن الأعداد الموجودة فى بطاقتين مثلاً فى وقت واحد، وبالتالى فالبحث عن الموضوع أ (المتعلق بالمصطلح ب) قد يظهر لنا أنه يتم تمثيله بواسطة أربعة مصطلحات موحدة، أما الموضوع، (ب) فيتم تمثيله بواسطة عشرة مصطلحات موحدة، ومعنى ذلك أن الباحث لابد أن يقارن (٤×١٠ بطاقات مستقلة). وواضح أنه إذا كان ذلك ممكناً، فهو عمل فيه كثير من الصعوبة، ويستغرق كثيراً من الوقت فضلاً عن صعوبة تطبيق الجبر البولينى Boolean Algebra ، وبالتالى فقد أصبح الحاسب الآلى هو الأداة لحل هذه المشكلات البحثية المتعلقة بتطويع أعداد كبيرة من المصطلحات غير المحكومة، ولكنها لا تستطيع بذاتها حل المشكلات الفكرية التى تنشأ عن عدم وجود المصطلحات المضبوطة أو المحكومة.

وفى مجال التعرف على تاريخ النظم المعتمدة على الحاسبات الآلية لاسترجاع المعلومات، يمكن التعرف على خطين رئيسين لهذا التطور ــ أحدهما ينبع من النظم الكبيرة التى قامت بها هيئات ضخمة كالمكتبة الوطنية الطبية أو وزارة الدفاع الأمريكية أو هيئة ناسا الفضائية. وقد قامت هذه الهيئات بتشغيل نظمها بناءً على المصطلحات المحكومة، أو التى تم وضعها للوثائق بواسطة المكشفين الأفراد لوصف هذه المصطلحات وتعدّ تمثيلاً موضوعياً لهذه الوثائق. أما الخط الثانى لتطوير نظم استرجاع المعلومات، فقد تم فى حقل القانون، وذلك بوضع النص الكامل فى شكل مقروء آلياً، ثم استخدام الحاسب الآلى للبحث عن الكلمات أو ارتباطات المصطلحات فى النص الواحد، وكان هذا النشاط سابقاً لتطوير المكانز، وسابقاً لظهور النظم الكبيرة التى تعتمد على التكشيف الإنسانى.

وعلى كل حال فقد كان المجال القانونى هو المجال الذى نبتت فيه الأساليب الحديثة لبحث النص الحر Legal Retrieval Systems ولقد أصبح نص الوثائق مختزناً فى الحاسب الآلى فى شكل مقروء آلياً أو كإنتاج جانبى لنشاطات النشر والبث، وبالتالى أصبح بحث النص عن طريق المستخلصات، فضلاً عن بحث النص الكامل مطبقاً لا فى المجال القانونى فحسب، بل فى المجالات الأخرى

كذلك، بل يمكننا القول بتزايد البحث فى الإنتاج الفكرى بواسطة كلمات النص، وليس بواسطة كلمات المصطلحات المضبوطة أو المحكومة، وإن كان ذلك مازال يحتاج إلى أدلة تجريبية قاطعة.

ويمكن تحقيق البحث فى النص بطريقتين:

الأولى: هى إدخال الكلمات فى الملف المقلوب الذى يظهر لنا الوثائق التى تتصل بهذه الكلمة (بل وعادة يعطينا الوضع المحدد لهذه الكلمة فى الوثيقة). الثانية: هى البحث فى النص بطريقة تتابعية كلمة بكلمة. دون استخدام الكشافات، وهذا هو الأسلوب المستخدم عادة فى تقديم خدمات البث الانتقائى للمعلومات (SDI) من قواعد المعلومات المقروءة آلياً، وذلك قبل انتشار نظم البحث على الخط المباشر، حيث اعتمدت هذه الطريقة على سمات المستفيدين المختزنة ومضاهاتها بالبحوث الجديدة الواردة لقاعدة المعلومات. أما اليوم فيمكن بواسطة الحاسبات بحث النص بسرعة بالغة، فعلى سبيل المثال فقد قام الباحث يو ورفاقه (yu et al, 1987) ببحث فى النص بمعدل ١٢,٥ مليون تمثيلة Character (حرف أو رقم) فى الثانية الواحدة، وهذا يوازى أو يساوى بحث حوالى خمسمائة صفحة فى الثانية الواحدة (٥٠٠ /ص/ث)، وعلى الرغم من أن هذا المدخل الأخير لا يختلف فكرياً عن مدخل بحث الكشاف المقلوب Inverted Index Approach ، إلا أن له إمكانات أفضل، حيث يتم البحث عن أجزاء من الكلمة وخصوصاً سلسلة التمثيلات أو اللواصق التى تحدث فى وسط أو نهاية الكلمة (Lancaster, 1991, 194)

ثالثاً: الحاجة إلى المصطلحات المحكومة:

هناك جوانب عديدة فى إعداد المصطلحات المحكومة، والإنتاج الفكرى حافل بالبحوث اللغوية فى هذا المجال، فقد أشار جوهانسن (Johansen, 1987) إلى أن إنشاء المصطلحات المحكومة يتطلب إنشاء علاقات بين الموضوعات ولابد من التمييز هنا بين العلاقات المعتمدة على "اللغويات" والعلاقات المعتمدة على تميز "اللغة" وقدم منهجاً للتعرف على بناء الموضوع عن طريق اختبار عناصره المكونة وعلاقات

أجزائه بعضها ببعض فضلاً عن تناول الموضوع من ناحية العلاقات التركيبية Syntactic فى لغات التكشيف.

ويعد استرجاع المعلومات فى الإنسانيات بصفة عامة أكثر صعوبة منه فى العلوم، وذلك لأن المصطلحات تعد أقل دقة؛ ويتناول الباحث ويبرلى Wiberley (1983 ذلك عن طريق دراسة المداخل المستخدمة فى الموسوعات والقواميس المشهورة فى مجال الإنسانيات، ويخلص من دراسته إلى أنه على الرغم من أن هناك بعض المصطلحات غير الدقيقة والمحددة imprecise إلا أن معظمها دقيق ويحتوى على أسماء أفراد أو أعمال إبداعية .. ولعل هذه النتائج تشير إلى أن الإتاحة الموضوعية فى الإنسانيات يمكن ألا تكون مشكلة جوهرية؛ وهناك الأهمية الكبرى لمعرفة اللغات نفسها، ذلك لأن هذه المعرفة تؤدى دوراً مهماً فى اختيار مصطلحات التكشيف والتحكم فى المكانز والإحالات التبادلية Cross-References فضلاً عن نقاط الإتاحة، ويشير الباحث إلى أن تطويع النص فى قواعد المعلومات العالمية يتطلب نوعاً من المعرفة بثلاث لغات Trilingualism وهذه تشمل معرفة بلغة المستفيد الوطنية، ومعرفة باللغة الأجنبية الصادر بها البحوث ومعرفة بلغة تشغيل النظام، وعدم القدرة على استغلال أى واحدة من هذه الجوانب الثلاثة سيؤدى إلى تقليل أداء النظام، ويخلص الباحث من ذلك إلى أهمية دراسة اللغات الأجنبية إذا أريد خدمة المستفيدين بطريقة أفضل.

وهناك مفهومان متعلقان بضبط المصطلحات وهما مفهوم السند الأدبى Literary Warrant ومفهوم سند المستفيد User Warrant فمصطلح التكشيف فى كتاب أو فى قائمة ضبط المصطلحات يدل على أن له سنداً أدبياً إذا كان هناك إنتاج فكرى كاف عن الموضوع وممثلاً بهذا المصطلح .. أما سند المستفيد فهو مفهوم مشابه، فيقال إن مصطلحا معيناً يعد سندًا Warranted إذا كان هذا المصطلح مستخدماً بطريقة منتظمة بواسطة المستفيدين للكشاف فى بحثهم عن المعلومات ... ومن الواضح أن سند المستفيد يعدّ مفتاحيًا بالنسبة لتصميم أداة ضبط المصطلح. كما يتطلب أن يكون المصمم مدركاً لما يحتاجه المستفيدون وعلى كل حال فالمصطلحات المضبوطة لها صفات مميزة كما يلى:

١- تمثل هذه المصطلحات البناء المفهومى العام للمجال الموضوعى، كما تقدم لنا مرشداً للمستفيد من الكشاف.

٢- تشتق المصطلحات من تلك المستخدمة أى أنها تعكس مصطلحات الإنتاج الفكرى وتعكس الاستخدام الفنى للمستفيد ... وعلى الرغم من أن هذه المصطلحات لا تستطيع أن تكون مخصصة Specific كما فى تكشيف اللغة الطبيعية، إلا أنها يجب أن تتيح إمكان الدقة للاسترجاع الجيد.

٣- تستخدم هذه المصطلحات المضبوطة عدداً مناسباً من الجمل سابقة التجهيز حتى تقلل من فرص الحصول على نتائج خاطئة False drops على قدر المستطاع.

٤- تزودنا بمصطلحات معيارية عن طريق التحكم فى المترادفات، وبالتالى تزيد من إمكان الانتظام، أى أنها تزودنا بآلية تسمح لنا باستخدام مصطلح واحد فقط من قائمة المصطلحات المشابهة لتكشيف مفهوم معين.

٥- يمكن أن تعرف لنا المصطلحات الغامضة، كما يمكن أن تميز بين مصطلحات الجناس Homographs (تماثل فى اللفظ واختلاف فى المعنى).

٦- عن طريق الإحالات يمكن أن تظهر لنا العلاقات الأفقية والرأسية بين المصطلحات.

وعلى الرغم من الإمكانات المحتملة الكبيرة لاستخدام اللغة الطبيعية فى الحاسب الآلى بواسطة المستفيدين إلا أن كليفيلند (Cleveland, 1990, 82) يذهب إلى أننا فى المرحلة الراهنة يجب أن ندخل فى حسباننا المستفيد الذى يفيد من المصطلحات المضبوطة أو المحكومة أو أنه يفيد من نظام لغة طبيعية محدود. ذلك لأن المصطلحات المحكومة ستعالج المترادفات وتربط بين المصطلحات المتقاربة دلالياً، وبالتالى تقلل عملية تخمين الكلمة الصالحة لدى المستفيد.

توليد المصطلحات وزيادة كفاءة التعبير والنمو :

هناك طريقتان لتكوين وبناء المصطلحات، الأول ما يمكن تسميته بالمصطلحات التطويرية evolutionary vocabulary وهذه تتكون كنتيجة لتوقف وضع المصطلحات الحرة عند نقطة معينة ثم محاولة وضع تنظيم معين لهذه المصطلحات .. أما الطريق الثانى فهو الذى يتم فيه توليد المصطلحات كنتيجة لدراسة متخصصة واتفاق بين الخبراء الذين يحددون ماذا ينبغى أن يدرج

من المصطلحات فى مجال معين من المعرفة .. وهذه الأخيرة تسمى المصطلحات الحصرية Enumerated vocabulary .

ونحن نلاحظ أن المصطلحات التطويرية تتكون من مواد خام يقدمها المكشفون، وبعد تكشيف عدد كاف من الوثائق يتم وضع المصطلحات فى ترتيب هجائى بغرض التحرير واتخاذ إجراءات القبول للمصطلح بعد تعريفه داخل مجال معين [مثال كلمة قاعدة (عسكرية) أو قاعدة (فى الرياضيات) أو قاعدة (فى الكيمياء) .. إلخ]، هذا بالإضافة إلى تحديد المرادفات ومعناها ثم وضع المرادف المفضل Preferred Term ويتم ضبط المصطلحات المترادفة بواسطة قاموس.

وليس هناك لغة قادرة على إعطاء مستوى الاتساع أو التحديد فى نقطة معينة، بمصطلح معين، وبالتالى يلجأ الباحث أو المكشف إلى توليفة من الواصفات لزيادة كفاءة التعبير عن المطلوب، وهذه تتم عادة بالتوليفات البولينية Boolean Combinations المستخدمة فى نظام الترابط اللاحق.

والآن لابد من الإشارة إلى إحدى خصائص نمو المصطلحات الكشفية ــ سواء الحرة أو المضبوطة ــ أى أنها دائمة التغيير وذات صفات ديناميكية .. وبالتالى فإن الإنتاج الفكرى فى موضوع معين يتغير، وبالتالى فإن لغة التكشيف التى تمثل محتويات الموضوع ستكون أيضاً ديناميكية متغيرة؛ والمصطلحات لا تتغير وحدها ولكن حجم الوثائق ينمو والموضوعات القديمة ستنقسم أو تنضم لتشكل مجالات جديدة [البيولوجيا والكيمياء تصبح كيمياء حيوية]، أى أن المصطلحات القديمة لم تعد صالحة أو كافية ومع ذلك فبقاؤها ضرورى لمسايرة التفكير والوضع القديم. أى أن لغة التكشيف تنمو بالضرورة مع ما يصحب ذلك من تعقيدات بنائية ... (Cleveland, D. 1990, 86-88).

المكنز :

تعد خطط التصنيف ورءوس الموضوعات والمكانز أدوات ضبط مرجعية، ولكن هناك بعض الصفات المميزة لكل منها، فعلى سبيل المثال فخطط التصنيف تكون عادة هرمية يصحبها كشافات هجائية ثانوية، بينما تكون المكانز عادة هجائية مع بناء هرمى مبنى بداخلها باستخدام الإحالات، والعلاقات الموجودة بين المصطلحات فى المكنز تكون عادة أكثر تخصيصاً، ومع ذلك فالواصفات فى

المكانز تكون عادة معتمدة على مصطلحات أخرى، ويتم ضمها مع مصطلحات أخرى، أما فى خطة التصنيف أو قائمة رؤوس الموضوعات فإن المصطلحات يمكن أن تقف بذاتها.

أى إن معظم مصطلحات رءوس الموضوعات مستقلة عن المصطلحات الأخرى، ويمكن لهذه المصطلحات أن تقف بذاتها، بينما مصطلحات المكنز يمكن أن تعتمد على مصطلحات أخرى؛ أى إنها تتعاضد مع بعضها حتى تكون أكثر تخصيصاً، وبهذا يكون لدينا نظم سابقة الترابط (كالتصنيف ورءوس الموضوعات) ولغات لاحقة الترابط (كالواصفات واللغات الطبيعية).

وإذا كان المكنز أساساً أداة ضبط لنظم التكشيف لاحق الترابط، بينما تعد قائمة رءوس الموضوعات ضابطة للنظم الهجائية السابقة الترابط أساساً، إلا أن درجة التمييز والفرق بينهما أصبح طفيفاً.

فالمكنز يضبط المصطلحات بطرق مختلفة فهو يحدد درجة تخصيص اللغة بتزويدنا بالمصطلحات التى يمكن استخدامها أو التى لا يمكن استخدامها، وبالتالى تقليص حجم المصطلحات، وكلما ازداد التخصيص فى اللغة، ارتفع معدل الدقة Precision فى الوثائق المختارة.

بناء المكنز :

يمكن أن يتم هذا البناء بطريقتين أولهما من القمة إلى أسفل مع مدخل نظرى حصرى مثل بناء نظم التصنيف أو بناء من أسفل إلى أعلى باستخدام مدخل عملى أمبيريقى، حيث يبدأ المدخل العملى هذا بالإنتاج الفكرى والمستفيدين منه شاملاً بعض الطرق الكمية كعدّ عدد مرات تردد استخدام الكلمة فى المجال الموضوعى.

وإذا ما اختير المسلك الأول النظرى فلابد من توفير عدد من الأدوات المساعدة مثل:

١- خطط التصنيف ورءوس الموضوعات خصوصاً تلك الخطط التى تخدم مجموعات متخصصة.

٢- المراجعات Reviews حيث تزودنا بالمصطلحات الجارية الشاملة.

٣- المنفردات Monographs .

٤- الأدوات المرجعية الأساسية مثل كتب الحقائق والقواميس والموسوعات وغيرها من المواد المرجعية التى تتناول المصطلحات فى الحقل الموضوعى مع الحذر دائماً من المصطلحات المتقادمة. Obsolete .

ويقترح الخطوات العامة التالية لبنا المكنز :

١- تحديد الحقل الموضوعى ومجالات التركيز فيه والمجالات الهامشية.

٢- تحديد طبيعة الإنتاج الفكرى الذى يتم تكشيفه، هل هو الإنتاج الفكرى الأولى للدوريات أم يشمل الكتب والتقارير وأوراق المؤتمرات ... إلخ؟ وهل الاهتمام بالإنتاج الجارى أم الراجع؟ وما يستتبع ذلك من سياسة تغيير فى المكنز.

٣- تحديد المستفيدين واحتياجاتهم المعلوماتية، وهل سيقومون هم بالبحث أم سيقوم المهنى فى المعلومات بهذه المهمة؟

٤- تحديد بناء الملف وهل سيكون نظام سابق الترابط أم لاحق الترابط؟

٥- استشارة الكشافات والقواميس وغيرها من الأدوات المنشورة فى المجال الموضوعى للتعرف على المصطلحات الخام، وهذا النشاط لا يتضمن بالضرورة اختيار المصطلحات بل يتضمن زيادة فهم المصمم للمصطلحات وعلاقاتها الدلالية.

٦- تجميع المصطلحات.

٧- وضع العلاقات بين المصطلحات (مصطلح أوسع BT مصطلح أضيق NT مصطلح قريب RT) فضلاً عن مفهوم (استخدم لـ USE) الذى يشير إلى الواصفات المفضلة من المصطلحات غير المستخدمة مثال: القطط الفارسية.

USE استخدم القطط.

القطط السيامية

USE استخدام القطط

أما مصطلح Use For استخدم لـ فيتناول أساساً المترادفات. أو الأشكال المتغيرة من المصطلح المفضل، كما تستخدم لإرشاد المكشف إلى مصطلحات أكثر عمومية مثال:

القطط

UF استخدم لـ القطط الفارسية

القطط السيامية

أما بالنسبة للأشكال فهناك بعض القواعد العامة التالية:

١- يجب أن تكون الواصفات أسماء Nouns- مفردة أو جملة أسماء أو أسماء مع تحديدات بين أقواس.

٢- المصطلحات المتعددة الكلمات يمكن أن تكون سابقة الترابط أو تتكون من المصطلحات الموجودة لاحقة الترابطة.

٣- الشكل المفرد يستخدم عادة للكلمات الدالة على العمليات أو الصفات، والشكل الجمع يستخدم لفئات الناس الذين يقومون بالعمل مثال:

العمليات التسبيل التكشيف الفئات المعلمون الواعظون

٤- يجب أن تدخل المصطلحات المتعددة الكلمات فى ترتيب الكلمات الطبيعى مع الإحالات إلى الأشكال المقلوبة.

٥- يجب استخدام المختصرات إذا كانت معانيها معروفة لدى المستفيد.

كما ينبغى تحديث المكنز لاستيعاب المصطلحات الجديدة فى المجال، فضلاً عن إحلال مصطلحات مكان أخرى وإجراء التغييرات فى العلاقات البنائية للمصطلحات القديمة أى أن الخطوات التى اتبعت فى إعداد المكنز يجب الالتزام بها أيضاً فى التحديث.

العرض والتقييم : Display & Evaluation

إذا كان المكنز عادة قائمة هجائية إلا أن له فى الغالب بعض الترتيبات الثانوية كالترتيب التصنيفى أو الهرمى بمجموعات الواصفات أو أحياناً العرض البرميوتى أو الدوار؛ والنماذج المثالية للمكنز تشمل:

- الترتيب الهجائى مع الإحالات.
- الترتيب الهجائى مع فئات موضوعية عامة.
- الترتيب الهجائى مع العرض الهرمى.
- الترتيب الهجائى مع التصنيف الوجهى.
- الترتيب المصنف مع كشاف

ومزية العرض المزدوج الهجائى ثم الهرمى هو أنه عند اختيار مصطلح المدخل فإن المكنز سيظهر فى الحال المصطلحات الأعرض والأضيق مع إظهار العلاقات الدلالية الموجودة .. وليس ذلك أمراً معيارياً لأن المكنز يمكن أن يقسم إلى جزأين أساسيين هما العرض التصنيفى الذى يظهر الأشكال الهرمية التفصيلية، والجزء الثانى وهو أداة المدخل الهجائى .. ولعل المكنز الوجهى هو خير مثال على ذلك أى أن هذا المكنز الوجهى هو توليفة من خطة التصنيف الوجهى التقليدى مع المكنز الهجائى، بحيث يظهر كل مصطلح فى المكنز فى خطة التصنيف الوجهى مربوطاً برمز معين هو رقم التصنيف عادة ... والترتيب الهجائى يدلنا على المصطلح المخصص، وعند الدخول به فإن الترتيب الهرمى سيظهر مكان المصطلح. أما العرض التصنيفى فيستخدمه الباحث عندما يكون لديه موضوع عريض فى ذهنه .. وعلى كل حال فالمكنز الوجهى وسيلة ممتازة لدمج مزايا التصنيف الوجهى التفصيلى مع إمكانات المكنز فى ضبط اللغة.

وأخيراً؛ فتقييم المكنز يتم بطريقة ذاتية أى مدى رضا المستفيد والمكشف، على الرغم من وجود طريقة شبه كمية بالنسبة للاستدعاء والدقة Recall & Precision وعلى كل حال فالمكنز له معاييره المتمثلة فى إنشاء الرابطة بين المفاهيم والمصطلحات عن طريق تجميع كلمات اللغة الطبيعية فى مجموعات ثم اختيار المصطلح المفضل Preferred للتعبير عن مفهوم الكلمات كلها فضلاً عن أن التقييم يشمل مدى التغطية الموضوعية ومدى كفاية العلاقات البنائية بين المصطلحات ومدى كفاية عرض المكنز لكفالة التكشيف والبحث الجيد.

رابعاً ـ المقارنة بين اللغة الطبيعية ولغة المصطلحات المحكومة فى استرجاع المعلومات:

يمكن تتبع الاهتمام ببحث النص بالحاسب الآلى إلى أوائل الخمسينيات، ولكن الاهتمام بهذا الموضوع استمر فى الزيادة مع توافر النص (المستخلصات أو المواد الكاملة) الذى يمكن إتاحته على الخط المباشر.

ولابد هنا من التنويه إلى أن النص الممكن إتاحته على الخط يتكون من مواد قصيرة كمقالات الدوريات والمستخلصات وبراءات الاختراع .. ونظراً لانخفاض التكاليف السريع (بسبب توافر الأقراص المكنزة مثلاً) فقد أصبح من الممكن اختزان وتطويع النص لمواد أكثر شمولاً ... ولكن إلى أى مدى سنكون قادرين على بحث النص الكامل .. للكتاب مثلاً؟

تشير العديد من الدراسات (Lancaster, 1989; P.55) إلى أنه بمقارنة نتائج البحوث فى القواعد الببليوجرافية المكشفة بواسطة الإنسان، بنتائج البحوث فى قواعد البيانات كاملة النص؛ أن الأخيرة تقدم لنا استدعاء أكبر greater recall ولكن دقة أقل reduced precision ويشير العديد من الباحثين إلى أن ذلك يعود إلى:

أ – طول التسجيلة أو المستخلص.

ب – تقليص عدد الواصفات المخصصة.

وقد يؤدى بحث النص الكامل فى أحيان كثيرة إلى إظهار مزيد من التفاصيل أما البحث بواسطة المصطلحات المحكومة فيميل إلى إظهار مفاهيم واسعة .. أى أن البحث الشامل يتطلب استخدام أكثر من طريقة واحدة ..

وينبه بعض الباحثين إلى محاذير عند وجود قواعد بيانات النص الكامل بطريقة عادية وميسّرة، ذلك لأن مجرد وجود النص الكامل قد يوهم بأنه قد يكون البديل الذى يوفر على المكشفين مهمتهم فى التكشيف الموضوعى.

وقد يرى المزج بين النص الكامل وأساليب البحث الموضوعى بديلاً مفضلاً للاسترجاع الأفضل.

وهناك بعض العوامل الرئيسة التى تؤثر على أداء نظم استرجاع المعلومات (كما هو الحال فى الشكل رقم (١) فيمكن أن نرى على الجانب الأيمن ثلاث تمثيلات للوثيقة من النص الحر (وهى العنوان واثنان من المستخلصات مختلفى الطول). أما على الجانب الأيسر ــ فهناك مجموعتان من المصطلحات الكشفية (التى تغطى الجانب الموضوعى بطريقة انتقائية أو شاملة)، والمصطلحات تعتمد على مكنز الأمم المتحدة UNBIS، ولعل أحد هذه العوامل الرئيسة التى تؤثر على أداء نظم استرجاع المعلومات هو عدد نقاط الوصول المتوافرة، ومن الواضح أن المستخلص الموسع يقدم لنا نقاط الوصول أكثر من المستخلص المختصر، وهذا بدوره يقدم نقاط وصول أكثر من العنوان، وقياساً على ذلك فإن التكشيف الموسع يقدم لنا نقاط وصول تفوق التكشيف الانتقائى فبحث النص على العنوان وحده ــ ربما يسترجع لنا المادة التى يعد موضوعها المحورى مركزاً فى العنوان وهكذا عند زيادة نقاط البحث، حيث يمكن التعرف على جوانب إضافية.

- وهناك جانب آخر فى عملية الاسترجاع هو طول التسجيلة نفسها، وليس نوع المصطلحات المستخدمة.
- ويمكن تحليل هذا الوضع ورؤيته بطريقة أكثر عمقاً عندما نختبر قاعدة معلومات تحتوى على آلاف المواد المكشفة بطريقة شاملة (كما هو الحال فى المثال الموجود فى الشكل رقم (١) فمثل هذه القاعدة توفر لنا استدعاء Recall أعلى من قاعدة بيانات توفر لنا الوصول من خلال العناوين فقط، وكذلك فإن قاعدة المعلومات التى تحتوى على مستخلصات موسعة ستوفر لنا درجة استدعاء أعلى من ذلك الذى يعتمد على التكشيف الانتقائى أو حتى التكشيف الشامل، وهذه المقارنة لا تتصل بمقارنة اللغة الطبيعية بلغة المصطلحات المحكومة فى استرجاع المعلومات، ولكنها تتعلق فقط بطول التسجيلة القابلة للبحث.

وخلاصة هذا كله أن المستخلصات ستزودنا بنقاط وصول أكثر من مجموعة الواصفات التى يضعها مكشف بعينه، كما أن النص الكامل للوثيقة سيؤدى بنا إلى

نقاط وصول أكثر أيضاً، وبالتالى نحن نتوقع أن قاعدة معلومات النص الحر ستقدم لنا استدعاء أكبر من ذلك الذى يعتمد على التكشيف الإنسانى.

- وهناك عامل آخر يؤثر على أداء نظام الاسترجاع وهو التخصيص Specificity، ذلك بالنسبة لوصف الجوانب الموضوعية للوثيقة، ويلاحظ هنا أن مصطلحات مكنز الأمم المتحدة UNBIS هى مصطلحات مخصصة للغاية لوصف معظم جوانب الموضوعات التى تظهر فى الشكل رقم (١)، ومع ذلك فإن النص الحر يقدم لنا درجة تخصيص أكبر.

الشكل رقم (١)

تأثير طول التسجيلة على الاسترجاعية

العنوان :	**تكشيف (انتقائى) :**
مسح الرأى العام الوطنى لاتجاهات الولايات المتحدة عن الشرق الأوسط.	- الرأى العام. - المسوحات الهاتفية.
مستخلص (مختصر):	- الولايات المتحدة الأمريكية.
يقدم مسح تم عن طريق الهاتف عام ١٩٨٥م وجهات النظر بالنسبة لبعض القضايا مثل: المعونة الأمريكية لإسرائيل ومصر؛ وهل تتخذ أمريكا موقفاً مواليا لإسرائيل أم للدول العربية أو موقفاً حيادياً؟ وهل يجب على منظمة التحرير الفلسطينية المشاركة فى مؤتمر للسلام، وهل تعدّ الدولة الفلسطينية المستقلة متطلب سابق للسلام؟	- الاتجاهات. - الشرق الأوسط. **تكشيف (شمولى):** - الرأى العام. - المسوحات الهاتفية. - الولايات المتحدة الأمريكية.
مستخلص (موسع) :	- الاتجاهات.
تم القيام بمقابلات هاتفية عام ١٩٨٥ مع عدد(٦٥٥) أمريكياً وكانت العينة احتمالية، وقد تم الحصول على إجابات للأسئلة التالية: هل إنشاء دولة فلسطين أساس للسلام؟؛ هل يجب تقليل المعونة الأمريكية لكل من إسرائيل ومصر؟؛ هل يجب مشاركة أمريكا فى مؤتمر للسلام بحيث يضم المؤتمر منظمة التحرير الفلسطينية؟؛ هل يجب على أمريكا عدم الوقوف إلى جانب إسرائيل أو إلى جانب الدول	- الشرق الأوسط. - إسرائيل. - مصر. - الدول العربية. - فلسطين.

العربية بل عليها أن تحتفظ بعلاقات صداقة مع الاثنين؟... ولقد تم التعبير عن الآراء بواسطة عدد من قادة الشرق الأوسط الرئيسين (حسين، عرفات، بيريز، مبارك، فهد، الأسد) خصوصاً بالنسبة لجهودهم من أجل السلام، فضلاً عن محاولة التعرف من المستجيبين إذا كان لديهم معلومات كافية عن الجماعات الوطنية المختلفة بالمنطقة	- منظمة التحرير. - مؤتمرات السلام. - السلام. - الدولة الفلسطينية. - المعونات الخارجية. - الزعماء السياسيون.

(Reference: Lancaster, 1991, P.7)

حيث يوفر لنا استرجاع أسماء قادة الشرق الأوسط مثلاً، بينما التكشيف يسمح لنا فقط بالبحث على مستوى القادة السياسيين. ومعنى هذا كله أنه كلما زادت نقاط الوصول Access points التى تساعدنا على الاسترجاع كان الاستدعاء Recall أعلى، وفى الوقت نفسه فستقل الدقة precision، والسبب فى ذلك ببساطة أنه كلما زادت نقاط الوصول، فإن بعض هذه النقاط سيتصل بجوانب ثانوية للوثيقة، ونتيجة لذلك فإن طالب السؤال الذى يتلقى الاستجابة عن مادة فى الشكل رقم (١) ، وبالنسبة لبحث عن القائد الفلسطينى عرفات، سيصل إلى نتيجة مفادها أن هذه المادة لا فائدة منها لأنها تتعلق بالقائد السياسى عرفات بطريقة مختصرة غير وافية.

وهناك عيب آخر يأتى من زيادة نقاط الوصول، وهى إمكان وجود بعض العلاقات الملتبسة Spurious وقد تكون هذه العلاقات فى شكلين:

أ – الارتباطات الزائفة False association .

ب- علاقات مصطلحات غير صحيحة. ويمكن أن يعكس لنا الشكل رقم (١) إمكانات عديدة فى هذا الشأن، فالمستخلص الموسع قد يؤدى إلى استرجاع مادة من خلال الذى تم فى مقابلة هاتفية مع قادة الشرق الأوسط (أو أى من هؤلاء القادة المذكورين). كما أن التكشيف الموسع قد يؤدى إلى استرجاعها فى بحث عن القادة السياسيين بالولايات المتحدة الأمريكية. فمثل هذه العلاقات والارتباطات هى بالضرورة ارتباطات أدت إلى استرجاع لا علاقة له بالوثيقة

(المقابلة الهاتفية لم تكن تتصل مباشرة بالقادة السياسيين، كما أن الولايات المتحدة الأمريكية ليست متصلة مباشرة بالقادة السياسيين).

وقد تظهر هذه العلاقات غير الدالة فى التكشيف الانتقائى أو حتى عند استخدام العنوان، وفى هذه الحالة فإن المصطلحات التالية (الشرق الأوسط ـ الاتجاهات ـ الولايات المتحدة الأمريكية) مصطلحات مترابطة فيما بينها، ولكن العلاقة غامضة.

والتكشيف الإنسانى هو عملية فكرية ذاتية، أى أن التميز الشخصى قد يدخل فى العملية، وبالتالى سيضيف المكشف موضوعات لا داعى لها أحياناً، كما قد يستخدم مصطلحا لا يمثل الموضوع تمثيلاً سليماً، كما أن المكشف قد لا يظهر العلاقات ذات الأهمية لمستفيدين بعينهم، ولعل النص الكامل قد يجنبنا هذا النوع من المشكلات، ولقد اهتدى الباحث (Horty 1962) إلى هذه الحقيقة منذ أكثر من عشرين عاماً حين قال:

عندما يستخدم النص الكامل للوثائق كأساس لنظام الاسترجاع، فإن سؤال الباحث لن يكون فى إجابته محدوداً بالطريقة التى يتم بها تكشيف الوثائق، فقد يتطرق الأمر إلى بعض الموضوعات المستقبلة، وقد يكون ذلك على قدر كبير من الأهمية، وهذا لا يتوقعه المكشف فى معظم الأحيان، وبالتالى فالاستغناء نهائياً عن الكشاف، والبحث المباشر فى النص الأصلى قد يعطى نتائج أفضل. وهناك جانب آخر لابد من الإشارة إليه وهو أن هذه المشكلة الخاصة بالمصطلحات غير المناسبة فى وثيقة واحدة تؤدى إلى علاقات زائفة فى الاسترجاع، هذه المشكلة تتضاعف وتتزايد مع تعدد الوثائق ومع تجميعها لتشكيل قاعدة بيانات ضخمة.

والربط الدلالى بين المصطلحات المتعلقة ببعضها Semantically سيساعد الباحث على تحديد جميع المصطلحات التى يحتاجها للوصول إلى بحث كامل.

- وهناك عامل إضافة مصطلحات جديدة تدخل فى العناوين أو المستخلصات، هذه المصطلحات لن تظهر فى المكانز وقوائم المصطلحات المحكومة إلا بعد مدة طويلة، وبالتالى فإن اللغة الطبيعية تعد ذات أهمية بالنسبة للموضوعات الجديدة وبالتالى ستحسن كل من الاستدعاء والدقة، وأخيراً فإن المصطلحات المحكومة

مفضلة لدى اختصاصى المعلومات الذى سيصبح بعد مدة قادراً على السيطرة على لغة البحث، بينما ستكون اللغة الطبيعية عادة هى المفضلة بالنسبة للمختص الموضوعى.

لعل المناقشة السابقة قد حددت بعض الخصائص المميزة للنص الحر وللتكشيف الإنسانى باستخدام المصطلحات المحكومة، وعلاقة ذلك بكل من الاستدعاء والدقة، وتتضح هذه العلاقات ويتم تلخيصها فى الشكل رقم (٢).

وواضح من هذا التحليل السابق أن الأمر ليس يسيراً، فهناك بعض العوامل التى تؤيد المصطلحات المحكومة، بينما يؤيد بعضها الآخر النص الحر، فتخصيص كلمات النص يميل إلى تحسين عامل الدقة، ولكن ذلك يجعل الوصول إلى استدعاء عالٍ أمراً عسيراً، على الأقل بالنسبة لبحوث المفاهيم العريضة، بينما يؤدى طول النص إلى تحسين الاستدعاء، ولكنه يقلل الدقة فى الوقت نفسه. ومع ذلك فلا نستطيع تفضيل هذا على ذلك أو على العكس لأن التفضيل هذا يعتمد على نوع البحث. فالبحث الذى يتضمن المفاهيم العريضة يفضل معه المصطلحات المحكومة، أما البحث الذى يتطلب التخصيص العالى (خصوصاً ذلك الذى يتضمن أسماء أفراد أو مؤسسات ... إلخ) سيحبذ النص الحر، أما البحث الشامل الحقيقى عن موضوع معين (مثلاً جميع المصادر الممكنة عن دواء معين)، هذا البحث يحبّذ استخدام النص الكامل، وأخيراً فالبحث الانتقائى (الذى يتطلب المواد الأكثر أهمية)، فهو يحبّذ التكشيف بالمصطلحات الحكومة.

ويمكن بصفة عامة أن يقال إن البحث باللغة الطبيعية يمكن أن يؤدى إلى نتائج أفضل بالنسبة للمفاهيم المحددة وليس بالنسبة للمفاهيم العامة (Fugmannl 1985) أما الباحث ديبوا (Dubois 1987) فيذهب إلى أن أحد مزايا النص الحر، هو إمكان استيعاب المصطلحات الجديدة. أما الباحث بيريز (Perez 1982) فيقرر أن المصطلحات المحكومة قد تؤدى إلى عدم الوصول إلى الدقة Precision، بينما لا يفقدنا النص الحر الموضوعات التخصصية.

كما ينبغى أن يؤخذ فى الحسبان جانب التكاليف عند مقارنة اللغة الطبيعية بالمصطلحات المحكومة فتكاليف التجهيز الفكرى الإنسانى تزيد بصفة مستمرة وبمعدلات عالية، وذلك عند مقارنتها بتكاليف التجهيز الآلى، كما أن التكشيف

بالمصطلحات المحكومة يتطلب عمالة كثيفة، فضلاً عن أن إنشاء ومتابعة المصطلحات المحكومة يعد أمراً مكلفاً، ومن جهة أخرى فنصوص المقالات والبحوث تصبح على مرّ الأيام متاحة بتكاليف زهيدة فى الشكل المقروء آلياً، وذلك كناتج إضافى لمناشط النشر والبث.

ومن هنا أصبح من الأمور الطبيعية أن يقيّم مدير خدمات المعلومات الموقف بعناية لتحديد مدى ميزات التكشيف بالمصطلحات المحكومة، وهل تعد التكاليف الإضافية المذكورة سابقاً مبرراً كافياً لاستمرار استخدامها؟

فالاستغناء عن التكشيف الذى يقوم به الإنسان باستخدام المصطلحات المحكومة، سيؤدى غالباً إلى تخفيض تكاليف المدخلات، ومع ذلك فإن ذلك يتم على حساب تكاليف كبيرة للمخرجات، على اعتبار أنه سيكون هنا عبء فكرى متزايد على المستفيدين من قواعد المعلومات، ولعل هناك بعض العوامل الأخرى التى ستؤثر على قرار المخرجات السابقة، ومن بينها حجم الوثائق والبحوث وتكاليف التكشيف والبحث ودرجة الأهمية التى يمكن أن نردها لنتائج البحث.

الشكل رقم (٢) المزايا والعيوب للنص الحر مقارنة بالمصطلحات المحكومة

* العوامل التى تساعد على الاستدعاء	أثر هذا النوع من التمثيل
أ – طول التسجيلة (عدد نقاط الوصول)	معظم تمثيلات النصوص الحرة ستكون عادة أطول من المجموعة المعينة من مصطلحات التكشيف. وهذا من شأنه أن يؤدى إلى تحسين الاستدعاء لكنه يقلل الدقة (وهناك زيادة فى حالات العلاقات الزائفة).
ب – التكرار غير المرغوب فيه (الحشو) Redundancy	هذه الخاصية ستكون عادة أكبر فى النص الحر، وبالتالى ستحسن من إمكان وجود مادة معينة، ومع ذلك فإن الطرق المختلفة الكثيرة التى يمكن التعبير بها عن موضوع معين فى قاعدة بيانات نصيّة ضخمة يجعل من العسير الوصول إلى استدعاء عالٍ.
ج – وجود مفاهيم ومصطلحات عريضة	وهذه تظهر عادة فى تمثيل المصطلحات المحكومة.
د–ربط المصطلحات دلالياً Semantically	وهذا واضح فى المصطلحات المبنية بناء سليماً (BT/NT/RT)
* العوامل التى تساعد على الدقة:	

- التخصيصية Specificity	النص الحر يكون عادة أكثر تخصيصاً، أى أنه يؤدى إلى الدقة المطلوبة، ومع ذلك فتنوع الطرق التى يتم التعبير بها عن المفاهيم يجعل من العسير جداً الوصول إلى استدعاء عالٍ فى بحوث المفاهيم العريضة، وفى مثل هذه البحوث فالمصطلحات العريضة المحكومة نسبياً ستكون المفضلة.
٨- عوامل تؤثر على كل من الاستدعاء والدقة:	
أ - المواكبة Currency	ستكون تمثيلات النص الحر دائماً أكثر مواكبة، ذلك لأن العثور على موضوع جديد فى نظام المصطلحات المحكومة يتطلب من الباحث تجريب عدد من المصطلحات الأخرى (وبالتالى تقليل الدقة) ومع ذلك فسوف لا يعثر على كل شيء عن هذا الموضوع (أى تقليل الاستدعاء).
ب - الألفة Familiarity	اختصاصى المعلومات يكون أكثر ألفة بالمصطلحات المحكومة ويستخدمها بطريقة أكثر فاعلية من غيره، أما المستفيد النهائى، فسيكون أكثر ألفة باللغة الطبيعية الموجودة فى الوثائق فى مجاله الموضوعى المحدّد.

بعض الدراسات السابقة :

لقد كانت المقارنات التى تمت فى الستينيات والسبعينيات بالنسبة لبحث النص الحر وبحث المصطلحات المحكومة تشير إلى أن النص الحر قد يكون أداؤه مثل المصطلحات المحكومة، إن لم يكن أفضل منها، ومع ذلك فيجب أن نذكر أن هذه الدراسات قد تمت على ملفات صغيرة، أى أنها كانت دراسات تجريبية، وليست خدمات معلومات حقيقية فى ظل نظام تشغيل طبيعى، ومنذ الثمانينيات تمت بعض الدراسات بالاستعانة بقواعد المعلومات ذات الحجم الأكبر، أو أن هذه الدراسات تناولت الخدمات الفعلية العاملة.

- من بين هذه الدراسات دراسة ماركى وزملائه (Markey et al. 1980) حيث قام بتحليل بحوث الكلمات المحكومة والنص الحر، وذلك بالنسبة للبحث على الخط المباشر لقاعدة بيانات إيريك Eric، كما قام ماركى وزملاؤه بمقارنة اختبارات بحث على الخط المباشر للنص الحر والمصطلحات المحكومة باستخدام ستة موضوعات، وكانت النتيجة أن النص الحر يقدم لنا استدعاءً

أعلى، أما المصطلحات المحكومة فتقدم لنا دقة أعلى، ولكن هذه التجارب كما يقول lancaster لم تحتوِ على تفصيل تقييم صلاحية أو تحكم فى استراتيجيات البحث أو غيرها.

- وهنا بعض الدراسات التى تمت فى حقل القانون، حيث قام الباحث كوكو (COCO 1984) بمقارنة نتائج البحث المعتمد على النصوص وحدها بتلك الناتجة عن النص، علاوة على إضافات التحرير، إلا أنه تبين أن قاعدتى البيانات اللتين قام بمقارنتهما لم يكونا متماثلتين بالنسبة لهذه المقارنة.
- هناك دراسة قام بها وانجر وزملاؤه وتمت على عدد (٥٣٥) بحثاً فى الـ Medline، وقام بها (١٩١) باحثاً مختلفاً، وكانت النتيجة متوسط استدعاء ٢٣ % بينما وصلت الدقة إلى ٦٧% (Wanger et al, 1980).
- ومن الدراسات الجيدة التى تمت لمقارنة النص الكامل والتكشيف المحكوم تلك التى قام بها تينوبير (Tenopir 1984) وكانت نتائجها كما يلى:

	النص الكامل	المستخلصات	المصطلحات المحكومة
عدد الوثائق المستوجعة (المتوسط)	١٧,٨	٢,٤	٣,١
عدد الوثائق المسترجعة الصالحة (المتوسط)	٣,٥	١,٠	١,٢
الاستدعاء	٧٣,٩	١٩,٣	٢٨,٠
الدقة	١٨,٠	٣٥,٦	٣٤
التكاليف لكل بحث	٢٠,٥٧ دولار	٤,٩٥ دولار	٥,٣٢ دولار
التكاليف لكل مادة صالحة مسترجعة	٧,٨٦ دولار	٣,٨٩ دولار	٣,٥٤ دولار

ويعلق لانكستر Lancaster على هذه الدراسة بأن أرقام التكاليف لا يمكن أخذها بجدية نظراً لأن تكاليف تينوبير (Tenopir 1984) قد شملت تكاليف الحصول على نسخ كاملة من الوثائق للتعرف على درجة الصلاحية، بينما لا يحدث ذلك فى الحياة الطبيعية إلا نادراً (حيث يتم حكم المستفيدين على أساس العناوين أو المستخلصات التى تصلهم على الخط المباشر أو بالأقراص المكتنزة). ولعل أهم النتائج التى وصلت إليها الباحثة تينوبير Tenopir هى أن بحوث المصطلحات المحكومة استرجعت بعض المواد التى لم يتم استرجاعها بالنص الكامل والعكس صحيح إلا أن هناك حاجة إلى الاثنين معاً.

إن مراجعة الإنتاج الفكرى هنا قد ركزت على الدراسات التى تقارن أداء قواعد المعلومات ذات النصوص الحرة بقواعد المعلومات المكشفة بواسطة المصطلحات المحكومة عندما يتم البحث بتوليفات المصطلحات طبقاً للجبر البولينى. ومن الواضح من هذه المراجعات الحماس المبكر لبحث اللغة الطبيعية، إلا أن هذا الحماس قد قل على مر السنين نظراً للمشكلات التى واجهها الدارسون وإن كانت الدراسات مازالت مستمرة لحل هذه المشكلات، ومن الأمور التى يجب الإشارة إليها أن البحث فى قواعد المعلومات التى تحتوى على آلاف المواد، هذا البحث إذا أدى إلى مستويات ضعيفة فى الدقة لا يعدّ أمراً مقبولاً، نظراً لعدد المواد المسترجعة، كما أنه من العسير الوصول إلى استدعاء عالٍ عند مستوى مناسب للدقة.

ومن المهم المتميز بين **مصطلح النص الحر والنص الكامل**، فالنتائج التى تم التوصيل إليها بالنسبة لقواعد معلومات النص الكامل لا يمكن أن تنسحب على قواعد المعلومات التى تحتوى على أقل من النص الكامل (يحتوى مثلاً على المستخلصات) فالنص الكامل يجب أن يعطينا استدعاءً أكبر بدقة أقل، وذلك عند مقارنته بقاعدة بيانات تحتوى على نص أقل من الكامل، وظهر ذلك فى التجارب التى قامت بها الباحثة تينوبير Tenoper ويلاحظ هنا أن النتائج فى واقع الأمر ترتبط بطول التسجيلة وليس بمقارنة اللغة الطبيعية بالمصطلحات المحكومة.

وخلاصة ذلك كله أن اللغة الطبيعية والمصطلحات المحكومة لها ميزاتها ولها عيوبها، فتسجيلات النصوص الحرة تميل إلى أن تكون أطول من غيرها، وبالتالى تزودنا بنقاط وصول أكثر، وستحتوى بعض المصطلحات الأكبر تخصيصاً أو

الأكثر حداثة من تلك المصطلحات المحكومة، أما المصطلحات المحكومة من جانب آخر، فهى تفرض عملية الانتظام أى تمثيل المحتوى الموضوعى بين الوثائق، وتمدنا بمصطلحات المفاهيم العريضة، التى لا تكون موجودة فى النص، وعن طريق التركيب الهرمى والإحالات تقدم للمستفيد معاونة إيجابية فى تحديد المصطلحات البحثية المناسبة.

ومن بين النتائج التى يمكن تقريرها هنا أن نظام الاسترجاع المثالى سيشمل بعض المصطلحات المحكومة فضلاً عن بعض النصوص الحرة، حيث يتم بكل منها استرجاع بعض المواد الصالحة.

التحكم اللاحق للمصطلحات :

دعا بعض المؤلفين إلى إمكان تحسين بحث اللغة الطبيعية عن طريق مكنز البحث أو مصطلحات التحكم اللاحق (Lancaster, 1986). ويذهب لانكستر إلى أن نظام مصطلحات التحكم اللاحق يمكن أن يقدم لنا جميع ميزات اللغة الطبيعية مع بعض مميزات مصطلحات التحكم القبلى، ومثل هذا النظام يمكن أن يكون أداؤه أفضل من نظام يعتمد على مصطلحات التحكم السابق.

وإحدى مميزات اللغة الطبيعية هى أنها قاعدة معلومات مستقلة، وبالتالى فإن جدولاً لـــ"خطوط الطيران سيكون مطبقاً فى جميع قواعد المعلومات باللغة الإنجليزية، ومن هنا يمكن أن نرى مكنزاً باللغة الطبيعية مطبقاً فى مئات من قواعد المعلومات.

ولعل أقرب الأدوات إلى مصطلحات التحكم اللاحق قاعدة تيرم TERM المستخدمة بواسطة خدمات الاسترجاع الببليوجرافى (BRS) التى قام بوصفها ناب (Knapp, 1983)، وهذه القاعدة هى قاعدة معلومات للجداول تمثل المفاهيم التى تشمل كلاً من المصطلحات المحكومة ومصطلحات النص الحر وهى المطلوبة للقيام بالبحوث فى العديد من قواعد المعلومات الاجتماعية والسلوكية، وفيما يلى نموذج لهذا الجدول:

TI	مناطق الفقر Poverty Areas
ER	+/ الفقر. مناطق Poverty Areas

ME	مناطق. الفقر Poverty Areas
PS	(انظر أيضاً: حى الأقليات GHETTO) مناطق. الفقر
SO	انظر: حى قذر للفقراء Slum حى الأقليات Appalachia GHETTO
EN	حى قذر للفقراء Slums
FT	مناطق الفقراء . Skid Row. Bowery. Slum. Inner City
	Poor Neighborhoods. Milieu of Poverty. Depressed Areas. Slums Ghettos. Ghettos. Ghetto, Ghettoes. Low Income Areas.

فالعنوان (TI) الدال على الجدول هو مناطق الفقر Poverty Areas، ويستخدم هذا المصطلح لاسترجاع المواد عن هذا الموضوع فى قاعدة بيانات إريك (ER) Eric وفى قواعد المعلومات المكشفة بواسطة قائمة رءوس الموضوعات الطبية (ME) وفى قاعدة علم النفس (PS) (حيث يوجد بها مصطلح مرتبط قريب هو GHETTOS) وفى المستخلصات الاجتماعية (So) وهناك مصطلحات أخرى ممكنة مثل Slum, Ghetts وهناك مصطلح أضيق فى قاعدة إريك Eric وهو (EN) أى Slums .. وأخيراً توجد قائمة تفصيلية بمصطلحات النص الحر (FT) القريبة والمفيدة والقيام بالبحث عن هذا الموضوع فى أية قاعدة معلومات باللغة بالإنجليزية .. ولسوء الحظ فإن قاعدة البيانات هذه ــ كما يقول لانكستر ــ لم تعد قائمة الآن.

ويستطرد لانكستر مؤيداً لوجهة نظره الخاصة بإمكان بناء مصطلحات التحكم اللاحق Post- Controlled Vocabulary فى أى حقل معين عن طريق الجهد الفكرى الإنسانى وذلك بالطريقة نفسها التى يبنى بها المكنز التقليدى، ويتم تيسير هذه المهمة عن طريق التطويع الآلى للكلمات التى تحدث فى قواعد المعلومات ذات الصلة وبالتالى يمكن أن تشتق derive مستويات عديدة من "الارتباطات الاحصائية" التى تساعد فى تنمية المكنز باستمرار.

خاتمة :

أثبتت النظم التى تستغنى عن التحكم فى المصطلحات التقليدية والتكشيف الإنسانى أنها نظم تستطيع أن تستمر، وقد فعلت ذلك خلال أكثر من ثلاثين عاما مضت. ومع ذلك فهذه النظم تواجه مشكلات عند ضرورة القيام ببحوث "للمفاهيم" العريضة، ذلك لأن هناك مميزات محددة للغة الطبيعية، إلا أن إثراءها المناسب (الاستخدام المحدود للتكشيف أو تطوير أدوات بحثية مساعدة) يمكن أن يعمل على تحسين فاعلية نظم اللغة الطبيعية.

أضف إلى ذلك ... أنه كلما ازدادت هذه القواعد حجماً وأصبح النص الكامل متاحاً للبحث، فسيصبح من الضرورى استخدام نظم تقوم بترتيب المواد المسترجعة عن طريق "الصلة المحتملة" Probable relevance وليس ببساطة تقسيم قواعد المعلومات إلى "مسترجعة" و"غير مسترجعة" (Maron, 1988).

خامساً – البحوث المستقبلة فى كل من اللغة الطبيعية والمكانز (اللغة المحكومة):

١/٥ تطورات مستقبلة فى معالجة اللغة الطبيعية وتجهزها:

تخضع أساليب المعالجة الآلية للغة الطبيعية للتطوير المستمر منذ بداية اختراع الحاسب الآلى، و قد أثبتت هذه الأساليب الفنية فى السنوات الأخيرة قوتها وكفاءتها بحيث تؤدى إلى منتجات تجارية فى العديد من المجالات ... وتطبيقات هذه الأساليب متوافرة فى مجالات الترجمة الآلية وفى تعامل الحاسب الآلى باللغة الطبيعية interfaces فضلاً عن التحليل الأسلوبى للنص وغيرها من التطبيقات.

وقد قام الباحث سميتون (Smeaton, A. 1992) وهو عالم فى الحاسبات الآلية بدراسة لمراجعة التطورات الحديثة فى استخدام مستويات التحليل اللغوى المعجمى Lexical والتركيبى Syntactic والدلالى Semantic والمخاطبى discourse فى التكشيف الآلى أو شبه الآلى للنص وفى استرجاع النص وفى استخلاص النص وفى توليد المكنز من جسد النص Text corpus، وأخيراً فى الاسترجاع المفهومى للمعلومات.

وقد قام الباحث سميتون وزملاؤه بدراسات فى مجال تطبيقات التحليل التركيبى syntactic analysis لمضاهاة وترتيب الجمل باستخدام التمثيلات البنائية للنصوص structured representation وعلى كل حال فستحاول هذه الدراسة التعرف على توقعات المستقبل بالنسبة لزيادة فاعلية ونوعية الاسترجاع المعلوماتى عن طريق المعالجات الآلية الحديثة للغة الطبيعية، ذلك لأن بحوث استرجاع المعلومات قد تطورت كثيراً خلال الأعوام الثلاثين السابقة باستخدام المداخل المعتمدة على الإحصاء والكلمات المفتاحية بطرق حديثة وذلك لمعالجة النص وتجهيزه بطريقة بنائية مثمرة.

والمعلومات كيان مجرد يمكن أن يوجد فى أشكال وأوعية عديدة، ولكن المعلومات تصبح شيئاً مفيداً إذا كانت متاحة للشخص المناسب فى الوقت المناسب ... ومع فيضان المعلومات وتدفقها المستمر أصبح الوصول للمعلومات المفيدة مهمة حاسمة.

لقد نشأت اللغة الطبيعية أساساً كآلية اتصال منطوقة، وتطور الاتصال المنطوق إلى اللغة المكتوبة لم يغير من شكل اللغة كثيراً، على الرغم من اختلاف اللغة المكتوبة عن المنطوقة بالنسبة للمصطلحات والبناء النحوى الذى نستخدمه .. وحتى اللغة المكتوبة فهى تختلف فى الأسلوب، فالوثائق الفنية تحتوى على كلام محكم محدد، حيث توجد عادة الجمل المعقدة لتحمل المعلومات الفنية الصعبة .. والجمل فى هذه الحالة واضحة معبرة فى طبيعتها.

أما القطع الصحفية كمقالات الجريدة اليومية فتحتوى عادة على جمل أقصر وأكثر سهولة وبساطة فى قراءتها ... أما القصص النثرية فى الكتب والروايات فيمكن أن تكون معقدة نسبياً ولكنها تقف فى منتصف الطريق بين الصحيفة والوثيقة الفنية وذلك بالنسبة لتعقد اللغة.

وتستخدم اللغة الرسمية المحكمة جداً والعسيرة فى القراءة فى الوثائق القانونية كالعقود والمعاهدات، وأخيراً فرسالات البريد الإلكترونى يمكن أن لا تتبع قواعد نحوية متعارف عليها، ويمكن أن يكون معظمها مختصرات وربما لا تحتوى على كلمات كاملة أبداً.

وخلاصة هذا كله أن مصطلح اللغة الطبيعية يمكن أن يشير إلى عدد كبير من الأنواع والأشكال وذلك طبقاً للتطبيق والاستخدام.

وإذا كان لدى المستفيد احتياجات معلوماتية غامضة، فإنه سيعبر عن هذه الاحتياجات باللغة الطبيعية أو بتوليفة كلمات مفتاحية بالجبر البولينى، ولكن هذا التعبير سيكون غير دقيق أيضاً، وسيطلب المستفيد الوصول إلى المعلومات التى قد تكون هى نفسها مكوّدة بطريقة غير دقيقة وفى لغة طبيعية ولكنها غير محددة. ولا تستطيع الكلمات المفتاحية معالجة الخصائص التالية للغة الطبيعية:

٥/١/١- كلمات مختلفة يمكن أن تستخدم لتحمل المعنى نفسه مثل "ألم المعدة بعد الأكل" "حرقان المعدة بعد الطعام" يلاحظ هنا قيام الكاتب بترجمة الجملتين الإنجليزيتين التاليتين:

"Stomach pain after eating" "Post Prandial abdominal discomfort".

٥/١/٢- الكلمات نفسها تستخدم لتؤدى معنيين مثل ستاير لحجب النور حسب المطلوب Venetian blinds والفينسيين العمى blind venetians أى استخدام الكلمة كاسم أو كصفة.

٥/١/٣- الناس المختلفون قد يكون لهم تصور مختلف عن المفهوم الواحد نفسه مثل "الحادثة" "والحادثة المؤسفة" قد يكون وصف الشيء نفسه ولكن فى موقف الدفاع أو الاتهام.

٥/١/٤- الكلمات نفسها لها معانٍ مختلفة فى مجالات مختلفة ... كلمة حاد Sharp قد تقيس شدة الألم فى الطب أو نوعية آلة القطع عند الجزار أو الجناينى.

وبالتالى يرى الباحث سميتون أن الوصول إلى المعلومات المعبر عنها باللغة الطبيعية يمكن أن يتم بنجاح فقط عن طريق معالجة أو تجهيز اللغة الفعلية للنص وليس فقط كلمات السائل التى حدثت .. ومع ذلك يظل الغموض وسوء التفسير قائماً بالنسبة للغة الطبيعية .. وهذا يحدث على جميع مستويات معالجة أو تجهيز اللغة الطبيعية. وبالتالى يجعل التجهيز الآلى لها عسيراً للغاية ولكنه ليس مستحيلاً.

وتعرف اللغويات المحسبة Computational linguistics بأنها ذلك المجال الذى يهتم بدراسة وتطوير النظم المحسبة للقيام بمعالجة أو تجهيز اللغة الطبيعية الآلية .. وهذه الدراسة لا تحاول تحديد قواعد أو مسار عالمى واحد للغة الطبيعية كما هو الحال مع اللغويات النظرية، ولكنها ـــ أى اللغويات المحسبة تهتم بتطوير الإجراءات اللازمة لمعالجة معظم حالات اللغة الطبيعية بالتلاؤم مع فشل عملية المعالجة أو التجهيز الذى قد يحدث بين وقت وآخر. وهدف اللغويات المحسبة هو وضع النظم الفعلية التى يمكن تطبيقها مثل الترجمة الآلية والتفاعل بين الإنسان والآلة Man-machine interfaces .

٢/٥- نظرة عامة على تجهيز اللغة الطبيعية أو معالجتها :

تنقسم عملية بناء النظم المعقدة التى يحتاجها تجهيز اللغة الطبيعية إلى نشاطين أساسين عادة، هما النشاط المستقل والنشاط المتعاون حيث يعملان على مستويات مختلفة من فهم اللغة. وقد كان التقسيم الأصلى بين المستويات واضحاً تماماً فى الماضى، ولكن هذا التمييز أصبح اليوم غير واضح بل ضبابى...

ومستويات تجهيز اللغة التى نهتم بها لأغراض استرجاع المعلومات هى المستويات المعجمية Lexical والتركيبية Syntactic والدلالية Semantic والمخاطبية Discourse (Gardar, G. 1989) (Grishman, R., 1986) .

وحتى يمكننا تجهيز إحدى جمل اللغة، يجب عزل أو تحديد عناصر أو مميزات تلك اللغة؛ ويعمل التجهيز المعجمى للغة الطبيعية على مستوى الكلمة المفردة، ويتضمن تطابق الكلمات وتحديد فئاتها النحوية حتى يمكن عمل التحليل اللغوى ذى المستويات الأعلى .. وهذا يتطلب عادة البحث فى القاموس أو المعجم عن قائمة الكلمات المعروفة وتصريفها المختلف morphological variants مثل الجمع للأسماء واسم الفاعل للأفعال .. إلخ Participles for verbs . والتحليل المورفولوجى أو الصرفى يتضمن تقسيم الكلمة إلى مكوناتها، وبالتالى فالكلمة الإنجليزية Covering ستقسم إلى Cover,ing والتجهيز المعجمى سيحدد نوع هذه الكلمة كفعل مضارع Present Participle للفعل Cover أو اسم مفرد singular noun.

وفى الأحوال المثالية فالمعالجة المعجمية ستحدد "شكل قاعدى واحد one base form" لكل كلمة وكذلك تاج تركيبى واحد one syntactic tag كذلك، ولكن ذلك لا يحدث عادة فى اللغة الإنجليزية، ذلك لأن العديد من الأسماء يمكن أن تقوم هى نفسها مقام الفعل، كما أن معظم الأسماء تنشأ عند جمعها بإضافة حرف (s) بالطريقة نفسها كشكل الفعل المفرد للشخص الثالث .. وهذه المشكلات يتم معالجتها أو حلها فى التحليل التركيبى (أو الإعراب).

ويلاحظ أن بحوث التحليل التركيبى للغة الطبيعية تهتم أساساً بقواعد النحو والصرف ذات التغطية الواسعة مع وضع استراتيجيات إعراب ذات كفاءة عالية، وقد أثبتت اللغة الطبيعية صعوبة كبيرة فى تحقيق ذلك نظراً لوجود استثناءات كثيرة.

وهناك المركبات الاسمية عندما يستخدم الاسم أو الأسماء كإحدى المقيدات النحوية modifier لاسم آخر، مكونة لبناء تركيبى كما هو الحال فى الجملة التالية Computer performance evaluation فكلمة Performance وهى اسم تقيد كلمة evaluation وهى اسم آخر، وكلمة Computer اسم يقيد الكلمتين evaluation أو Perfrmance ولكننا لا نعلم على أيهما يتم التقييد، وهنا أيضاً يوجد الغموض أو الالتباس البنائى.

وعلى الرغم من الجوانب السلبية فهناك العديد من المميزات المهمة لتجهيز اللغة على المستوى التركيبى بما فى ذلك تحديد بناء الجملة، كما أن قواعد النحو والتركيب عامة فى طبيعتها، كما أن المفاهيم كالكلمات ذات مستوى تجريدى Abstract؛ ويخلص الباحث سميتون من هذا العرض العام إلى أن تجهيز اللغة الطبيعية يستخدم الآن فى استرجاع المعلومات، وقد أدى لظهور ما يسمى باسترجاع المعلومات المفهومى Conceptual Information وهو مجال من مجالات الذكاء الاصطناعى (zani, G., 1983) واسترجاع المعلومات المفهومى يختلف عن استرجاع المعلومات التقليدى، ذلك لأن المستفيد فى النظام التقليدى تقدم له قائمة نصوص أما فى النظام المفهومى فلا تقدم له قائمة نصوص، ولا تقدم له مجرد إشارات مرجعية، ولكن تقدم

له المعلومات نفسها مباشرة وبالتالى فالنظام المفهومى أكثر صعوبة وتعقيداً بالنسبة لتجهيز اللغة.

بعض نتائج بحوث سميتون وزملائه:

عالجت ورقة البحث المستويات الأربعة للغة والتى لها اتصال باسترجاع المعلومات ... فعلى المستوى المعجمى lexical level يبدو أن القواميس المقروءة آلياً ستقدم لنا إمكانات مهمة للتكشيف وتمثيل النصوص، ولكن يبقى هناك ضرورة لكثير من العمل التجريبى حتى يمكن الوصول إلى كفاءة استخدام القواميس المقروءة آلياً ... والتكشيف بمعانى الكلمات Word senses باستخدام القواميس المقروءة آلياً سيؤدى بنا إلى استرجاع للنص يكون أكثر فاعلية من مجرد التكشيف بجذع الكلمة word stem، كما أن كمية البحوث الضخمة فى استرجاع المعلومات إحصائياً خلال العقود السابقة يمكن استخدامها لتمثيل معانى الكلمات وليس جذور أو جذوع الكلمات.

ولعل أبسط التطبيقات فى تجهيز اللغة الطبيعية لاسترجاع المعلومات قد تمت على مستوى الكلمة وعن طريق التكشيف المعتمد على كلمات مأخوذة من قائمة استناد أو مشتقة من النص ... وهناك بديل لذلك عن طريق الشكل القاعدى للكلمات base forms words من معجم أو قاموس. ولكن بناء مثل هذا القاموس يعد أمراً مكلفاً ولم تكن له إلا تحسينات هامشية على الشكل الآلى وبالتالى فلم تتحقق هذه الفكرة.

ومع ذلك فقد كان للتحليل اللغوى على المستوى المعجمى أهميته الحديثة مع زيادة توافر القواميس المقروءة آلياً (MRD) بما فى ذلك قاموس أكسفورد الإنجليزى على أقراص مليزرة CD-ROM ... وهذا القاموس (MRD) يحتوى على تعريف لكل معنى sense أو تفسير للكلمة بما فى ذلك الفئة Syntactic Category، ووصف للنص مختصر وربما بعض المعلومات الدلالية semantic .

وعلى كل حال فتجارب استرجاع المعلومات التى تضمنت التكشيف بواسطة معانى الكلمات word senses وباستخدام القواميس المقروءة آلياً لم تكن ناجحة بدرجة كبيرة بالنسبة لفاعلية الاسترجاع .. وبالتالى فقد غير الباحثون من

اتجاههم، حيث أصبح من غير الضرورى تحديد المعنى الصحيح الواحد single crrect sense للكلمة، ولكن يكفى استبعاد المعانى غير المحتملة ثم وزن المعانى بدرجة عالية .. ولقد قام كل من كروفتش وكروفت بهذه التجارب مستخدمين وزن المصطلح إحصائيّاً ولكنهما لم يحصلا بعد على التحسينات المتوقعة فى نوعية الاسترجاع (Krovetz, R. 1989)، ومازالت تجاربهم وتجارب الآخرين على قدم وساق.

أما المداخل التركيبية Syntacical للتكشيف فقد كانت مجال بحوث عديدة حديثة .. ولكن أهم عيوب هذا الاتجاه هو أنه يتعامل مع نصوص على مستوى الجملة فقط ولكنه لا يتعامل مع قضايا مثل الأنفرة anaphora وهناك عيب خطير آخر مع استرجاع المعلومات المعتمد على الإحصاء، وهو أنه لم يلق اهتماماً لقضية المترادفات بين الكلمات أو بين المقاطع phrases وهناك محاولات لحل هذه المشكلة فتجهيز المعلومات المعتمد على الجانب الدلالى Semantic يمكن من تكشيف هذه المترادفات فى الإطار التمثيلى نفسهsame frame-based representation . ويفترض العديد من الباحثين أن الاسم أو الجمل الاسمية في اللغة هى المعبرة عن المحتوى وعن وحدات المعلومات، ولكن ذلك ليس صحيحاً دائماً خصوصاً بالنسبة للتمثيل الكامل للمعنى، وإن كانت الأسماء كشافاً جيداً لمحتوى النص بغرض استرجاع المعلومات التقليدى وهو المطلوب فى معظم الأحيان.

وهناك افتراض آخر كبير مفاده أن الذين يستخدمون المستوى التركيبى لتجهيز اللغة الطبيعية فى استرجاع المعلومات إنما يفعلون ذلك لأن التركيبى اللغوى Syntax أى البناء اللغوى يشرح لنا الدلالة أو المعنى؛ أى أن البناء اللغوى يتضمن المحتوى. وهذا غير صحيح فى أحوال عديدة خصوصاً إذا كان البناء والتركيب اللغوى نفسه غير واضح ويحتمل وجهات نظر ambiguous .

ومع ذلك ففى التطبيقات القريبة من الكمال فنحن فى حاجة إلى التجهيز الدلالى لاستبعاد مشكلات التركيب وللوصول إلى تمثيل أكثر عمقاً وأكثر ثراء.

وهناك تطور مهم للغاية فى اللغويات المحسبة Computational Linguistics خلال السنوات القليلة الماضية، وهو ظهور التجهيز اللغوى المعتمد على الإحصاء .. وقد قام براون (Brown, K. etal, 1990) بتحليل إحصائى على عدد ثلاثة ملايين

زوج من الجمل إحداها باللغة الإنجليزية والأخرى الترجمة الفرنسية .. ولقد أظهر التحليل احتمال تجاور الكلمات المعتمد على السياق Context وعلى وجود كلمات أخرى. ومن هذه الإحصاءات أنشيء نظام ترجمة آلي يحتوى على حوالى تسعة آلاف مصطلح، ووصلت نسبة نجاح الترجمة ٥٣% بالنسبة للجمل الكاملة حيث ترجمت إلى المعنى الصحيح .. وعلى الرغم من أن هذه التجارب كانت مكلفة للغاية من ناحية الحاسب الآلى، إلا أن هذا العمل يظهر أن الترجمة المعتمدة على الإحصاء ممكنة. أما تشرش (Church, K. 1988 فقد طور نظاماً لبناء قواعد نحوية احتمالية تستخدم حالياً لمحاولة تحديد الدلالات من الإحصاءات semantics from statistics وتعد هذه التطورات ذات أهمية بالغة لاسترجاع المعلومات نظراً لإمكان دمج التحليل اللغوى المعتمد على الإحصاء مع الطرق التى سبق تطويرها على الإحصاء. وهناك توقعات مستقبلة للتطوير فى هذا الاتجاه.

وعلى الرغم من التقدم الذى حدث فى استرجاع المعلومات التقليدى إلا أن نظام استرجاع المعلومات المثالى هو الذى يقدم لنا استرجاع المعلومات المفهومى Conceptual ولسوء الحظ؛ فإن ذلك يحتاج إلى تجهيز اللغة على المستوى الدلالى Semantic وهذا بدوره يحتاج إلى قاعدة ضخمة للمعرفة المتخصصة فى المجال؛ أى إن هذا العيب يجعل عملية التحليل اللغوى كلها وكذلك استرجاع المعلومات المفهومى محدودة بمجال ضيق محدود. فتجهيز اللغة على المستوى الدلالى لا يرقى لأبعاد حجم استرجاع المعلومات، ولكن لدينا من التجارب الكثير الذى يشير إلى إمكانات أكبر فى المستقبل.

وأخيراً؛ فالدور الذى تؤديه حالياً أساليب تجهيز اللغة الطبيعية بالنسبة لبحوث استرجاع المعلومات يعدّ دوراً أمبيريقياً ... فأساليب تجهيز اللغة الطبيعية يمكن اعتبارها كالصناديق السوداء أو الأدوات التى تساعدنا على تقديم تكشيف أفضل وأكثر ثراء بواسطة العبارات Phrases وليس بواسطة الكلمات، وهذه الأساليب تزودنا بترتيب مضاهاة لهذه العبارات ... إلخ. ولكن هذا الدور لا يعالج قضايا استرجاع المعلومات للمستفيدين اعتماداً على اللغة المستخدمة فى أسئلتهم أو فى النص.

٥/٣- بعض التجارب الجارية والمستقبلة على المكنز الآلى:

٥/٣/١- البناء الآلى للمكنز الكونى:

٥/٣/١/١- ملخص:

هذه تجربة قام بها عالم الحاسب الآلى كروتش Crouch للتعرف على جدوى بناء مكنز كونى اعتماداً على ما يسمى بنموذج قيمة استبعاد المصطلح والخوارزم التجميعى المناسب، وإذا كان هناك اعتراف منذ زمن بعيد بفوائد ومزايا المكنز المبنى بطريقة جيدة بالنسبة لنظام استرجاع المعلومات، فقد كانت هناك بحوث وتجارب سابقة تناولت بناء المكنز بوسائل يدوية ونصف آلية وآلية، وثبت من هذه التجارب أن توليد المكنز آلياً يعد مشكلة أكثر صعوبة من الطريقتين الأخيرتين، ويحاول الباحث كروتش (crouch, e.,J., 1990) التعرف على المداخل المبكرة والجارية الحديثة بالنسبة لبناء المكنز الآلى، ثم يصف لنا مدخلاً للتوليد الآلى للمكنز الكونى معتمداً فى ذلك على كل من نموذج قيمة استبعاد المصطلح Term discrimination value model الذى تناوله الباحثون سالتون Salton ويانج Yang ويو Yu وعلى الخوارزم التجميعى المناسب appropriate Clustering algorithm . وقد طبقت هذه الطريقة على مجموعتين من الوثائق .. وقد أثبتت لنا النتائج الأولية لتجارب كروتش Crouch أن هذه الطريقة والتى أدت إلى تحسين فى الأداء الاسترجاعى بحوالى (١٠-١٥%) وذلك فى مجموعة الاختبار .. هذه الطريقة مفيدة وتستحق المزيد من البحث والتجريب. أى أنه من الممكن بناء وتوليد المكنز آلياً، ولكن نتائج دراسة كروتش تحتاج إلى مزيد من الدراسة حتى يمكن أن تصل إلى مرحلة التعميم generalization .

٥/٣/١/٢ مقدمة :

تعتمد فعالية استرجاع المعلومات إلى حد كبير على مجموعة القواميس المتوافرة لاستخدامها فى النظام، وإذا كانت هناك قواميس متعددة الأنواع (إحصائية أو تركيبية لغوية أو هرمية المصطلح ... إلخ) إلا أن المكنز يحتل أهمية أساسية بالنسبة لكفاءة أداء النظام .. ويمثل بناء المكنز الآلى صعوبة خاصة مقارنة ببناء المكنز اليدوى أو النصف آلى.

وقد أطلق الباحث كروتش اسم المكنز الكونى Global thesaurus على المكنز التقليدى فى بيئة استرجاع المعلومات .. وقد استخدم مصطلح "كونى" ليميزه عن المكنز المحلى local thesaurus الذى قام بوصفه وتحليله كل من العطار وفرنكل (Attar, R; Fraenkel, A, 1981) ويرى كروتش أن المكنز الكونى ــ بعد بنائه ــ يستخدم لتكشيف كل من الوثائق والأسئلة Queries .. وبالمقارنة بالمكنز المحلى فإن الأخير يستخدم المعلومات التى يتم الحصول عليها من الوثائق المسترجعة استجابة لسؤال معين وذلك لتعديل هذا السؤال .. وهذه بدورها يتم إدخالها ثانية إلى نظام الاسترجاع للمعالجة .. ونتيجة لذلك فإن المكنز الكونى يتم بناؤه قبل عملية التكشيف ويستخدم لتكشيف كل من الوثائق والأسئلة بينما يتم بناء المكنز المحلى أثناء عملية معالجة السؤال ويستخدم لتعديل هذا السؤال فقط.

وتحاول هذه الدراسة التى بين أيدينا اختبار مختلف طرق بناء المكانز بما فى ذلك بناء المكنز آلياً اعتماداً على ما يسمى بنموذج قيمة استبعاد المصطلح Term discrimination واعتماداً على الخوارزم التجميعى المناسب appropriate Clusterings Algorism ويعتمد نموذج قيمة استبعاد المصطلح بدوره على نموذج حيز المتجهات Vector space الذى سيتم شرحه فيما يلى :

٣/١/٣/٥- نموذج حيز المتجهات:

يعدّ نموذج حيز المتجهات أحد النماذج الرئيسة فى استرجاع المعلومات، حيث يرى هذا النموذج كل وثيقة – فى مجموعة الوثائق – بوصفها تمثل كلمات فريدة متميزة .. ويمكن عدّ كل وثيقة كمصطلح متجه Term Vector وبالتالى تصبح المجموعة الكاملة للوثائق كحيز متجهات ذى بعد عددى يساوى (m) حيث تعد (m) عدد أنواع الكلمات فى المجموعة.. وفى نموذج حيز المتجهات، فإن متجه الوثيقة (DJ) يعد فيه رقم الوثيقة المعنية (J) ويتم تمثيلها بمجموعة من المصطلحات (Djk) حيث تتراوح قيمة K من (١) إلى (m) وحيث تمثل (DjK) عدد مرات (أو وزن) المصطلح (K) فى الوثيقة (J)، وعلى سبيل المثال، إذا كانت (djk) تساوى (صفر) فإن المصطلح (k) لا يظهر فى الوثيقة (DJ). أما الأسئلة – مثل الوثائق – فيتم تمثيلها بمصطلحات متجهة موزونة Weighted term .

ومما ينبغى الإشارة إليه أن نموذج المتجهات Vector model يعطى عدداً من المزايا تفوق تلك التى تحصل عليها من النموذج البولينى Boolean model الذى يستخدم عادة فى نظم الاسترجاع الببليوجرافى، ذلك لأن نموذج المتجهات ييسر جوانب عديدة منها إيجاد رتبة Ranking الوثائق المسترجعة، والتغذية المرتدة الصالحة، وتجميع ملفات الوثائق وبناء الأسئلة فضلاً عن وزن المصطلحات Term weighting .

٥/٣/١/٤- نموذج قيمة استبعاد المصطلح :

يتم فى نموذج قيمة استبعاد المصطلح تعيين أدوار محددة للمصطلحات المفردة ومصطلحات الجمل ومصطلحات الفئات Term Classes، وذلك بغرض تحليل المحتوى وتزويدنا بإطار يتم بداخله ترتيب كل مصطلح كشفى محتمل فى المجموعة، وذلك طبقاً لفائدته كمميز للوثيقة. (Salton, G.; Yang, C.S., Yu, Ct, 1975).

كما أن هذا النموذج يزودنا بتفسير مادى معقول لعملية التكشيف. ويذهب الباحث كروتش (Crouch, C, 1990, 631) إلى أن نموذج قيمة استبعاد المصطلح يقدم لنا معياراً لتكوين المكنز الكونى، حيث يتكون المكنز من مجموعة فئات المكنز Thesaurus Classes وتتكون فئة المكنز من مجموعة من المصطلحات أو أنواع الكلمات .. وبالتالى فحتى يمكننا استخدام المعيار المقترح لهذا النموذج يجب حساب قيمة استبعاد كل مصطلح فى المجموع (طبقاً لعدد مرات حدوثه) ثم ترتيب هذه المصطلحات فى فئات (جيد/ لا علاقة له/ ضعيف) حسب قيمة كل مصطلح. وقد قام كروتش بدراسة مقارنة بين المصطلحات والوثائق وانتهى من مقارناته إلى أن عدد مرات تكرار الوثيقة يمكن استخدامه كقيمة تمييز أو استبعاد تقريبية كذلك.

٥/٣/١/٥ الإجراءات اللازمة لبناء المكنز الكونى:

يذهب كروتش إلى أن هذه الإجراءات تتضمن ما يلى:

أ – يتم تجميع مجموعات الوثائق من خلال الخوارزم التجميعى المناسب.

ب- يتم استعراض الهرمية الناتجة من هذا التجميع، ثم يتم توليد الفئات Classes اعتماداً على معايير محددة مقدمة من لدن المستفيدين.

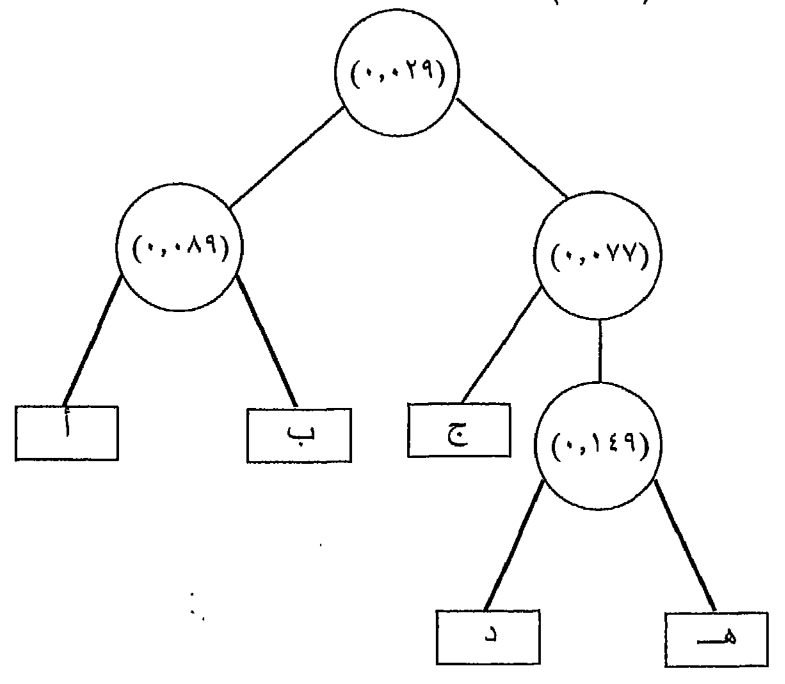
ج – يتم تكشيف الوثائق والأسئلة بواسطة فئات المكنز.

هذا وصفات فئات المكنز التى يتم توليدها فى الإجراء الثانى أعلاه تتحدد بواسطة المعايير التى يقدمها المستفيدون وهى كما يلى:

أ – القيمة المحققة Threshold value :

يمكن التعرف على هذه القيمة من المثال التالى:

إن تطبيق الخوارزم التجميعى الكامل الحلقات Complete-link clustering algorism يؤدى إلى هرمية تحتوى على أكثر التجمعات ترابطاً فى أسفل شجرة التجمع .. وهذه هى أوراق الشجرة فعلى سبيل المثال الشكل التالى، يظهر فيه المربعات (الأوراق) تمثل الوثائق والأعداد فى الدوائر تمثل المستويات التى تتجمع عندها الوثائق .. فمثلاً الوثيقتان أ،ب تتجمع عند القيمة المحققة (٠,٨٩)، ويلاحظ أن القيم المحققة تتناقص فى الاتجاه الأعلى حتى جذر الشجرة حيث القيمة المحققة تساوى (٠,٠٢٩).



ويلاحظ أن القيمة المحققة المطلوبة من لدن المستفيد، تحدد إلى درجة كبيرة الوثائق التى يمكن اختيار المصطلحات منها لإدخالها فى فئة المكنز .. ففى الشكل السابق تعدّ القيمة المحققة (٠,٠٩٠) تقبل فقط تجمع الوثيقتين (د،هـ) نظراً لأن المستوى الذى يتم فيه تجميع الوثقتين يجب أن يكون أكبر من (أومساوياً لـ) القيمة المحققة المطلوبة حتى يمكن أخذ هذا التجمع فى الحسبان.

ب- عدد الوثائق فى التجميع:

نظراً لأن كلا من (د،هـ)، (ج،د،هـ) تحقيق القيمة المحققة (٠,٠٧٥) فى الشكل السابق، فهناك معيار آخر مطلوب لتحديد عدد الوثائق التى يجب شمولها فى التجمع النهائى، هذا المعيار الذى يقدمه المستفيد هو عدد الوثائق فى التجمع، وفى المثال السابق فإن القيمة (٢) لهذا المعيار ستضمن لنا فقط التجمع (د،هـ)، بينما القيمة (٣) ستسمح بقبول التجمع (ج،د،هـ).

ج – العدد الأدنى لتكرار الوثائق Minimum document Frequency :

متى تم اختيار تجميعات الوثائق، تتكون فئات المكنز من المصطلحات ذات التكرار المنخفض الموجود ضمن هذه التجميعات، وقبل اختيار هذه المصطلحات، فإن العدد الأدنى لتكرار الوثائق الخاصة بكل مصطلح يجب أن تحدد. والقاعدة المرشدة فى الإجراء (٣) أعلاه يمكن أن تساعد فى وضع هذا المعيار.

د – طريقة تخصيص الفئة Specification of Class Formation method :

عندما تتحدد تجميعات الوثائق، ويتم تجميع المصطلحات منخفضة التكرارات من الوثائق المجمعة، فيمكن فى هذه الحالة توليد فئات المكنز ... وأبسط المداخل اللازمة لتحديد فئة المكنز هى تقاطع intersection جميع المصطلحات منخفضة التكرارات فى التجمع خصوصاً مع التقاطعات البسيطة.

٥/٤- توليد المكنز الآلى للنظام الإلكترونى :

٥/٤/١ ملخص :

قام بهذه التجربة أربعة من الباحثين ثلاثة منهم بجامعة أريزونا (قسم نظم المعلومات الإدارية) والرابع من جامعة ألينوى (قسم علم المكتبات والمعلومات)،

وقد تناولت هذه الدراسة مدخلاً خوارزمياً للتوليد الآلى للمكنز لنظام مجتمعى إلكترونى. وشملت الأساليب الفنية المستخدمة ترشيح المصطلحات Term filteering ، التكشيف الآلى وتحليل التجميعات Cluster analysis .. وكانت بوتقة الاختبار هى نظام مجتمع الديدان Worm Community System ويحتوى النظام على مكتبة شاملة للبيانات والإنتاج الفكرى المتخصص الذى يستخدم فى وقت الدراسة بواسطة علماء البيوليوجيا الـذين يدرسون دودة النيماتودا .. وقد احتوى مكنز النيماتودا الـذى يشمل عـدد (٢٧٠٩) أسماء باحثين، وعدد (٧٩٨) اسم جنس، وعدد (٢٠) طريقة تجريبية وعـدد (٤٣٠٢) واصف موضوعى .. وفى المتوسط كان لكل مصطلح حوالى (٩٠) مصطلحاً موزوناً قريباً weighted neighboring والذى يظهر لنا مفاهيم متعلقة.

وقد تم وضع المكنز كأداة بحثية على الخط المباشر .. وقام الباحثون باختبار مكنز النيماتودا فى تجربة تضم ستة باحثين ذوى درجات مختلفة من الخبرة والخلفية العلمية .. وقد أظهرت التجربة أن المكنز كان أداة ممتازة لتنبيه الذاكرة Memory Jogging وأن هذا المكنز قام بدعم عملية التعلم والتصفح الجيد.

وعلى الرغم من وجود بعض التشويش الواضح، إلا أن النظام كان مفيداً بالنسبة لاقتراح مفاهيم صالحة لأسئلة الباحثين، كما ساعد النظام فى تحسين استدعاء المفاهيم، ويمكن بالتصفح البسيط أن يصبح المكنز الآلى أداة مفيدة للبحث على الخط المباشر، كما يساعد الباحثين فى اكتشاف واستعراض نظام مجتمع إلكترونى معقد وديناميكى.

٢/٤/٥ بعض التوجهيات المستقبلة التى وضعها فريق البحث والمستوحاة من هذه الدراسة وهى كما يلى:

أ – التوليد المكنزى الإضافى:

يتم توليد المكنز فى الوقت الحاضر (عام ١٩٩٥م) بطريقة الدفع Batch mode (فى حوالى أربع ساعات) ونظراً لأن قاعدة بيانات نظام مجتمع الديدان (W C S) تتزايد بصفة مستمرة، فهناك حاجة إلى وضع طريقة فعالة للإضافة إلى المكنز وتحديثه .. وقرر الباحثون الأربعة أنهم يقومون حالياً بتطوير نسخة إضافات

للخوارزم التجميعى. وما سيجعل تحديث المكنز والإضافة إليه ممكناً هو اختزان بعض النتائج الوسيطة لتكرارات المصطلحات وبعض تكرارات الوثائق المقلوبة.

ب – معالجة تطبيقات أكبر مما تم بالتجربة:

لقد أظهرت لنا الدراسة الحالية إمكان اتباع أسلوب الحقل المستقل الآلى لإنشاء مكنز على الخط المباشر للحقول العلمية independent –automatic, domain approach to the creation of On-line Thesauri والبحوث الجديدة للفريق تحاول اكتشاف مجتمعات بحثية جديدة، ربما تحتوى على كميات أكبر من المعلومات بل تعانى من الفيض الهائل للمعلومات Information Overload ومن مشكلات المصطلحات فى استرجاع المعلومات (مثل قاعدة بيانات الإنتاج الفكرى للفيزياء) أى أنه باختيار مدخلهم البحثى فى مجتمعات أكبر، سيكونون أكثر قدرة على التحقق من أساليبهم البحثية وتعميمها.

ج– التقاط مرونة المفاهيم Capture the fluidity of concepts :

حتى الآن لم يشمل فريق البحث جوانب أخرى مثل البعد الزمنى للوثائق والمفاهيم فى التحليل .. ويعتقد فريق البحث أنهم عندما يتعرفون على زمن كل مفهوم وتحليل النشاطات المرتبطة به (متى ظهر لأول مرة؟ ومتى استخدم بدرجة كبيرة؟...) فسيتمكنون من إنشاء مكنز أكثر مرونة وأكثر دقة بالنسبة لعامل الزمن.

د – الاستشارة الآلية للمكنز:

الامتداد الطبيعى للبحث الجارى هو اختبار خوارزميات البحث العام المعتمدة على الذكاء الاصطناعى (Chen etel., 1993) وذلك بغرض استشارة المكنز آلياً .. وهذه الخوارزميات ستكون قادرة على معاونة الباحثين فى استعراض السعة المفهومية جميعها عن طريق اتباع الحلقات الأكثر صلة فى البداية ..وقد قام الفريق ببحوث فى هذا الاتجاه، ولكن مازالت هناك حاجة إلى التجريب المستمر لتطوير استشارات مكنزية آلية ضخمة لنظام مجتمع الديدان (W C S).

هـ – اتجاه السعة المفهومية اللازم لحل مشكلة المصطلحات A concept space approach to solving the vocabulary problem :

وأخيراً؛ يعتقد فريق البحث أنهم يقتربون من إيجاد حل لواحدة من أكثر المشكلات تحدياً فى استرجاع المعلومات وهى مشكلة المصطلحات.

ولمما كان الباحث فى حقل معين (مثل عالم البيولوجيا فى مجال الذبابة) يحتاج عادة لبحث الإنتاج الفكرى فى حقول قريبة (مثل بيولوجيا الديدان) مستخدماً فى ذلك مصطلحاته هو (مصطلحات محددة عن الذبابة)، فإن فريق البحث يحاول إنشاء سعة مفهومية أخرى خصوصاً بتحديد مدى التشابك overlap بين المصطلحات (مثلاً التعرف على مفاهيم البيولوجيا المشتركة) فى سعتين مفهوميتين: وهما السعة المفهومية للذبابة والسعة المفهومة لقاعدة بيانات الهدف وهى الديدان ... ومع تبنى عملية استشارة متعددة المفاهيم للمكنز؛ فإن مصطلحات الباحث المخصصة يمكن استخدامها لاستعراض السعتين المفهوميتين الأخريين، وفى الواقع دمج المصطلحات المخصصة فى سعة المفهوم للديدان (الهدف) ... وقد قام الفريق بالفعل بوضع مكنز الذبابة (Chen et al, 1994) وذلك كمرحلة أولى فى عملية التجريب الخاصة بتشابك واستعراض سعة المفاهيم لكل من الذبابة والديدان.

الفصل الرابع

التكشيف الآلى: دراسة للتطورات فى المجال خلال أربعين عاماً

تقديم :

لقد تعلم الإنسان منذ قرون أن يخترع وأن يستخدم الآلات لزيادة قدراته العضلية والمادية، ثم تعلم فى وقت متأخر، كيف يستخدم الآلات أيضاً لتزيد من قدراته العقلية كذلك، وكما كان الفأس فى الزمن القديم وسيلة الإنسان لتيسير تقطيع الأشجار ثم استخدامها فى أغراضه الحياتية واليومية، فإن الحاسب الآلى قد يسر على الإنسان فى الوقت الحاضر القيام بمهام معالجة وتجهيز البيانات لخدمة أغراضه المعاصرة المتمثلة فى بناء القوة فى عصر المعلومات، لقد دخلت الحاسبات فى أعمال المكتبات شأنها فى ذلك شأن أنشطة أخرى عديدة خارج المكتبات، كان ذلك بالنسبة للأعمال الروتينية أكثر منه بالنسبة للأعمال الفكرية، وأثبتت الحاسبات فاعليتها فى هذه الأعمال الروتينية بدرجة كبيرة، ولكن دخول الحاسبات فى مجال التحليل الموضوعى وبالذات فى التكشيف والاستخلاص، مازال يغلفه الكثير من الصعوبات، وإذا كانت نصوص الوثائق والأوعية المتعددة الأشكال بما تحتويه من رسومات، قد أمكن تحويلها إلى الشكل المقروء آلياً، فإن القواعد السليمة التى يمكن برمجتها بالحاسب الآلى، مازالت فى طور التجريب، بهدف الوصول إلى كشافات ومستخلصات آلية لا تختلف عما يقوم به العقل الإنسانى المبدع لدى المكشفين والمستخلصين المحترفين.

وتجدر الإشارة فى البداية إلى أن الباحثة ستركز فى هذه الدراسة على التطورات التى حدثت وتحدث فى هذا المجال بالنسبة للإنتاج الفكرى الأجنبى حتى عام ١٩٩٧م بما يتضمنه من بعض التجارب فى التكشيف الآلى للوثـائق العربية والنشورة باللغة الإنجليزية (*).

(*) هناك دراسات على جانب كبير من الأهمية عن استرجاع المعلومات والتكشيف الآلى صدرت ضمن الإنتاج الفكرى باللغة العربية ومن أهمها دراسات كتبها: حسين الهبائلى ـــ حشمت قاسم ـــ على الصوينع ـــ ناصر سويدان ـــ نبيل على ـــ محمد الأطرم ـــ وهانى عطية وغيرهم.

وستتناول الباحثة فى هذه الدراسة المحاور التالية:

أولاً: استرجاع المعلومات ومشكلات الطرق الآلية فى التكشيف.
ثانياً : المعالجة الإحصائية لكلمات النص: نشاط محورى فى التكشيف الآلى.
ثالثاً : التكشيف بالاستعانة بالحاسب الآلى.
رابعاً: التكشيف الإقتباسى الآلى والتكشيف التعيينى الآلى.
خامساً: تجارب فى التكشيف الآلى للوثائق العربية فى نظام استرجاع المعلومات.
سادساً : أشكال واتجاهات أخرى لتحسين التكشيف الآلى.

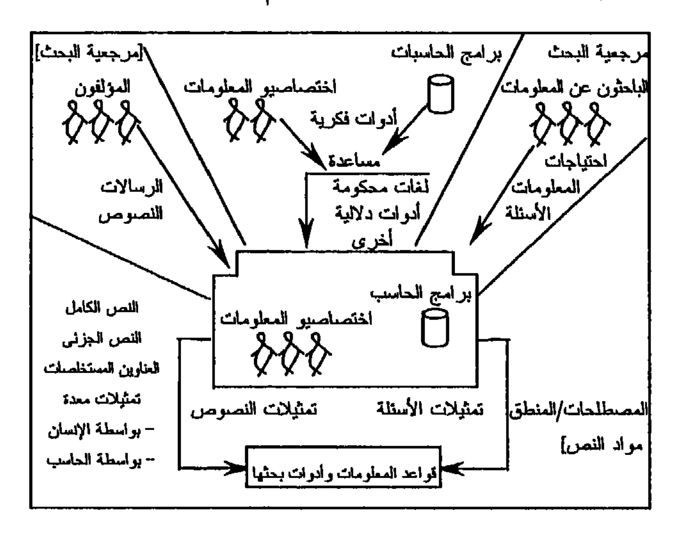
أولاً: استرجاع المعلومات ومشكلات الطرق الآلية فى التكشيف:

أ – استرجاع المعلومات وتطوره :

لقد تكونت خلال العقود الثلاثة الأخيرة مجموعة ضخمة من قواعد المعلومات على الخط المباشر، وتشير هذه الاتجاهات إلى مزيد من المجموعات الضخمة. ولكن ما تأثير ذلك على استرجاع المعلومات، إن أوضح تأثير هو زيادة عمق التكشيف نظراً لإمكانيات الحاسبات الهائلة فى التجهيز السريع لكميات ضخمة من المعلومات، وفى حالة توافر كل من الأشكال المطبوعة وعلى الخط المباشر، فستحتوى الأخيرة على واصفات أكثر تخصيصاً، وذلك لأن الكشافات المطبوعة المقيدة فى شكلها المادى، تكون محدودة بالنسبة لتغطيتها الشاملة. ومشكلة استرجاع المعلومات فى أساسها تتصل بمضاهاة احتياجات المستفيدين من المعلومات مع الرسائل التى أعدها المؤلفون وهذه يمكن أن تتم بطريقة غير مباشرة فمعظم الرسائل (ما يريد المؤلفون أن يوصلوه) تظهر على هيئة نصوص (بعضها على هيئة صور أو شكل سمعى أو غير ذلك من الأشكال غير النصية)، بينما تقدم احتياجات المعلومات كطلبات لبعض أنواع خدمة المعلومات، هذا وتنشىء خدمة المعلومات تمثيلات للنصوص وتختزن هذه فى قواعد البيانات، ثم تزودنا ببعض الأدوات التى تتيح بحث هذه التمثيلات، ويمكن أن تختزن قاعدة المعلومات على هيئة ورق أو صور مصغرة أو شكل إلكترونى. أما الأداة التى تسمح لنا بالبحث فقد تكون ببساطة ترتيب المداخل فى الفهرس البطاقى أو الكشاف المطبوع أو مجموعة برامج الحاسب، هذا وتنشىء خدمة المعلومات كذلك تمثيلات للطلبات، وتقوم بتجهيزها من قاعدة المعلومات مسترجعة تمثيلات النص التى تضاهى أو تقدم أفضل مضاهاة لتمثيلات الطلب.

هذا وقد تحتوى تمثيلات النصوص على النصوص الكاملة ذاتها (نسخ إلكترونى للكلمات المطبوعة على الورق)، أجزاء من النص أو بعض أشكال التمثيلات الإنسانية أو الآلية، أما تمثيلات الطلبات فستحتوى على مصطلحات أو مصطلحات مقدمة فى علاقات منطقية أو بيانات نصية أو مواد (والمقصود بالمواد هنا نظام يتيح للباحث إدخال تفاصيل مواد معروف سلفاً صلاحيتها، وذلك لطلب مواد تشبهها بطريقة أو بأخرى).

وما يساعدنا على بناء التمثيلات (من النصوص أو من طلبات المستفيدين) هناك بعض الأدوات الفكرية المتاحة وأكثرها وضوحاً هو المصطلحات المحكومة التقليدية، وهناك أيضاً المصطلحات لاحقة التحكم.



(Lancaster, F., 1991, p. 220)

هذا ويجب أن تكون مشكلات استرجاع المعلومات واضحة من الرسم السابق، فالنصوص قد لا تكون تمثيلات دقيقة للرسائل (هذه يمكن أن تكون مشكلة اتصالية محدودة، ولكننا لا نعتبرها عادة مشكلة استرجاع معلومات) وبنفس الطريقة فطلبات المستفيدين قد تكون تمثيلات دقيقة لاحتياجات المعلومات فقد لا تتطابق مرجعية السائل مع مرجعية اختصاصى المعلومات أو على الأصح مرجعية المؤلفين أنفسهم، فمشكلة استرجاع المعلومات إذن يمكن اعتبارها كمشكلة محاولة مضاهاة أقرب ما

يمكن من احتياجات المعلومات إلى أقرب ما يمكن من الرسائل (Lancaster, F., 1991, P. 220) وتشير الباحثة بيتز (Betes, 1986) إلى مشكلة استرجاع المعلومات على أنها مشكلة عسيرة، أى أكثر تعقيداً مما قد تبدو على السطح، وهى تشير إليها باعتبارها مشكلة احتمالية وأننا نركز الآن على جانب المخرجات من النشاط (احتياجات المعلومات ــ الطلبات ــ التمثيلات) أكثر مما نركز على جانب المدخلات (الرسالة ــ النص ــ التمثيلات) أى أن الافتراض الضمنى هو أن الجانب الخاص بالمخرجات هو أكثر صعوبة وتعقيداً.

ويمكن استخدام الحاسبات الآلية فى التكشيف والاستخلاص الآلى، فضلاً عن استخدامها فى عمليات أخرى تتضمن تكوين فئات الوثائق أو المصطلحات وفى وضع استراتيجيات البحث، وفى إنشاء شبكات التجمعات بين المصطلحات وكما يدلنا الشكل السابق تستطيع الحاسبات الآلية إلى حد ما أن تكون بديلة للإنسان فى مختلف الأنشطة المبينة تقريباً، وفى الوقت الحاضر لا تستطيع الحاسبات أن تولد رسائل أو احتياجات معلوماتية بطريقة مستقلة، إلا إذا تم برمجتها لتقوم بذلك عن طريق الإنسان نفسه، وإن كان من الممكن أن تقوم الحاسبات فى يوم ما بهذا العمل.

ويمكن إيجاز مميزات الحاسبات الآلية فيما يلى :

❁ تتبع الحاسبات التعليمات بطريقة آلية:

تعنى برمجة الحاسب إعداد قائمة تفصيلية من التعليمات والتى تدل الآلة على ما تقوم به فى سلسلة منطقية. والعمل الإنسانى الوحيد بعد إدخال البرنامج بالحاسب، هو مدخلات البيانات عند نقاط عديدة، ثم تقوم الآلة بإتمام المهام بطريقة آلية.

❁ الحاسبات تتميز بالدقة المتناهية:

عندما يقوم الحاسب بعملية جمع لملايين الأرقام فى ثوان معدودة، فمن غير المحتمل حدوث أى خطأ فى الحساب، وهذه الخاصية تجعل الحاسب ذا أهمية وفائدة بالغة فى جميع قطاعات التطبيقات المتصلة، ومع ذلك فالمبرمجون قد يخطئون، عندما يغذون الحاسب بتعليمات خاطئة.

❁ الحاسبات تواجه المشكلات العسيرة التي يتم تحليلها بكفاءة :

هناك عامل ثابت فى اختلال وظيفته نظام الحاسب وهو فشل المبرمجين فى الفهم الكامل للمشكلة، ذلك لأن دخول الحاسب فى هذه الحالة سيؤدى إلى تضخيم الأخطاء القديمة وليس حلها. كما ينبغى التأكيد على أن فهم المشكلة شيء، وفهمها بتفصيل وعمق شيء آخر، ذلك لأن الفهم التفصيلى العميق لازم للبرمجة السليمة والوصول إلى الحلول الممكنة. والآن هل يمكن إدخال التعليمات والبرامج للحاسب والتى تصلح لتكشيف واستخلاص التسجيلات؟ إن مهارة المبرمج الإنسان هى التى ستدفع بالقدرات المنطقية المحسبة للأمام لاجتياز هذا الطريق. (CLEVELAND, 1990, P. 222)

(ب) مشكلات الطرق الآلية فى التكشيف:

هناك العديد من الدراسات خاصة تلك التى قام بها جيرارد سالتون (SALTON, G., 1971) والتى تظهر لنا أن النظم الآلية تقوم على الأقل بما يقوم به المكشفون بالطرق اليدوية، ومن الاعتراضات الأولى على النظم الآلية بصفة عامة وعلى التكشيف بصفة خاصة هى أن "الحاسب يمكن أن يكشف الكلمات فقط ولكنه لا يكشف المفاهيم Concepts ولكن المفاهيم هذه لا يمكن أن يتم تمثيلها إلا بالكلمات، وما يعكسه هذا التفكير هو أن التكشيف بالتعيين Assigned Indexing (حيث تستخدم فيه المكانز ونظم التصنيف ورؤوس الموضوعات مثلاً) أفضل من التكشيف بالاشتقاق Derived Indexing (مثل كشاف كويك KWIC) ولكن البحوث المتعددة تتحدى هذا الاعتقاد وتنقضه. فهل كلمات المكشف أفضل من كلمات المؤلف والذى يفترض فيه أنه يعرف فى الموضوع أكثر مما يعرف المكشف وستقوم الباحثة بتفصيل ذلك فيما بعد. وليس معنى هذا الذى سبق أن التكشيف الآلى ليس له مشاكله. بل هناك مشاكل أساسية تتمثل أساساً فى أن التكشيف الآلى مازال فى مرحلة التجريب ولم يصل إلى التطبيق على نطاق واسع بعد وبالذات مع النظم الحقيقية الكبيرة.

هذا وعملية التكشيف التى يقوم بها المكشف عملية مكلفة وتتراوح جودتها بين التميز والرداءة، ومع النمو السريع للمعلومات، فقد زادت الفجوة الزمنية بين نشر ورقة البحث وتوافر الكشافات والمستخلصات لها، والحلول المقترحة لمواجهة هذه

المشكلة لم تعد كافية، فزيادة عدد المكشفين ليس حلاً مرضياً وقد لا يكون في الإمكان تحقيقه اقتصادياً، كما أن المكشفين المهنيين المؤهلين يصعب الحصول عليهم، ومن هنا ظهرت الحاجة ماسة إلى استخدام الطرق الآلية.

ومع ذلك ولسوء الحظ فلسنا قادرين حتى وقتنا الحاضر على كتابة البرامج التى تسمح للحاسب بفهم اللغة الطبيعية كما يفعل الإنسان، هذا وتكمن قوة الحاسب فى معالجة وتجهيز أدق التفاصيل (يقوم الإنسان بحل المشكلة المعقدة عن طريق تقسيمها إلى سلسلة طويلة من الخطوات التفصيلية ثم يتولى الحاسب الآلى المهمة بعد ذلك)، ذلك لأن الحاسب الآلى لا يقوم بالأحكام القيمية التى يقوم بها الإنسان، والمعروف أن الأحكام القيمية هذه تلعب دوراً رئيسياً فى إنشاء كشافات ومستخلصات ذات نوعية وجودة عالية.

ثانياً: المعالجة الإحصائية لكلمات النص: نشاط محورى فى التكشيف الآلى:

بدأت أهم الأعمال الجادة فى مجال التكشيف الآلى فى منتصف الخمسينيات عندما أعلن الباحث لوهن (Luhn, H., 1957) أن المصطلحات التى تحتويها الوثيقة أو أسئلة المستفيدين يمكن أن تستخدم لأغراض تحليل المحتوى وقال لوهن فيما قال ما يلى: إن الفكرة التى تتكرر مرتين على الأقل فى نفس الفقرة Paragraph يمكن اعتبارها فكرة رئيسية، والفكرة التى تحدث سابقة أو لاحقة مباشرة للفقرة يمكن اعتبارها فكرة رئيسية حتى لو ظهرت مرة واحدة فقط فى الفقرة موضع الدراسة، ويمكن بعد ذلك عمل رموز رئيسية Notation فى ترتيب معيارى معين. ولكن أفكار لوهن الأصلية لم تلق رواجاً أو اعترافاً لأسباب عديدة منها:

١- إلغاء العلاقات بين الكلمات.

٢- معيار عدد مرات تردد الكلمات لتحديد مدى أهميتها لم يثبت عليه دليل يمكن تعميمه فى ذلك الوقت وإن كانت هذه الطريقة الأخيرة قد طورت مع باحثين آخرين بعد ذلك لتمثل ركناً أساسياً يعتمد على الإحصاء واللغويات وتمثلت أهمية وفائدة المصطلحات بناء على عاملين أساسيين (Salton, G., 1975, p. 85) هما:

(أ) الدقة العالية: High Presision فيمكننا أن نتساءل هنا كيف يمكن لمصطلح معين أن يستخدم لاسترجاع بعض الوثائق ذات الصلة الكبيرة Very relevant وهذا يعطى قيمة لعدد مرات التردد العالية فى عدد قليل من الوثائق.

(ب) الاستدعاء العالى High recall وهنا يكون الاهتمام بعدد الوثائق ذات الصلة والتى يمكن استرجاعها بمصطلح معين وهذا يضع قيمة لعدد مرات الوثائق الكافية فى التوزيعات التكرارية frequency distributions وفى تقييمه لمنتجات التكشيف الآلى ذهب لانكستر عام ١٩٦٩ (Quated by Salton, G. 1975, P. 98) إلى أن الإمكانية الفنية والاقتصادية للتكشيف الموضوعى الآلى قد ظهرت بوضوح عن طريق تحقيق التكشيف السريع وإن كان متوسطاً فى جودته Mediocre وتستخدم نظم التكشيف الآلى حالياً فى عدة مؤسسات على مستوى إنتاجى، حيث تستخدم فى إعداد مطبوعات تكشيف منذ عام ١٩٦٠ ومن أمثلتها كشافات العناوين الدوارة [كويك أو كووك Kwic or Kwoc] وهو تجهيز بكلمات العنوان وحتى هذه المنتجات تشمل نظم تجهيز النص الكامل.

هذا وقد تمت دراسات تقيمية عديدة خلال الستينيات وأوائل السبعينيات ومعظم هذه الدراسات تناولت اختبارات نظريات التكشيف السائدة فى تلك الفترة فعلى سبيل المثال:

(أ) ضرورة استخدام التحليل التركيبى Syntactic analysis ويتمثل فى التجهيز والمعالجة الآلية لمستخلصات التعدين من أجل استرجاع المعلومات وهى البحوث التى تمت بجامعة كيس وسترن ريزرف ـ مركز بحوث التوثيق والاتصال.

(ب) الارتباطات الإحصائية للمصطلحات.

(ج) القواميس اليدوية والآلية.

(د) محاولة قياس فاعلية عدد كبير من لغات التكشيف وأهمها دراسات كرانفيلد وسمارت وميدلرز Medlars / Cranfield and Smart (*) ومن أهم النتائج التى انتهى إليها الباحثون بعد الدراسة والمقارنة هى أن لغة التكشيف البسيطة غير

(*) MEDLARS = Medical Literature Analysis and Retieval System.

المحكمة والتى تتضمن مصطلحات مفردة Single Terms تؤدى إلى أفضل أداء استرجاعى، بينما تؤدى المصطلحات المحكومة. والعبارات (مفاهيم بسيطة) (Phrases Simple concepts) إلى نتائج أسوأ (Salton, G., 1975, P. 101) وفى موضع آخر يذكر سالتون أيضاً بأن نتائج الاختبار (مقارنة نظام ميدلرز بنظام سمارت بالنسبة للاستدعاء والدقة) تدلنا مرة أخرى أن نظم تحليل النص الآلى البسيط نسبياً لا تؤدى فى بيئة استرجاع الوثائق إلى نتائج متدنية Inferior عن التكشيف المحكوم (Salton, G., 1975. p. 105) .

هذا وتعتمد أكثر الأساليب المنتشرة فى مجال التكشيف الآلى على عدد مرات تكرار الكلمات فى النص، أو عدد مرات حدوثها بالتلاحم مع غيرها من الكلمات أو عدد الكلمات التى تحدث بين مصطلحين وهكذا (Cleveland, 1990, 224). لقد نشر العالم جورج زيف Zipf كتابه عن "السلوك الإنسانى ومبدأ أقل الجهد) عام ١٩٤٩ حيث أثار الاهتمام لعامل سلوكى هام يتعلق بعدد مرات تكرار الكلمات فى نصوص اللغة الإنجليزية، وذهب زيف إلى أن اللغة محكومة بمبدأ أقل الجهد، بمعنى أنه على مدى فترة من الزمن، فإن اللغة ستجد أقل عدد من الكلمات فى نصوص مختلفة وانتهى إلى الملاحظة التالية: إذا ما تم ترتيب الكلمات فى نص معين بحيث تكون الكلمة الأكثر استخداماً فى البداية ولها رتبة رقم (١) وكانت الكلمة التالية فى كثرة الاستخدام تأتى فى الرتبة الثانية أى برقم (٢) وهكذا فسيكون لدينا ما يلى:

الرتبة × عدد مرات التكرار = رقم ثابت

Constant (RF = C)

ولكن إحدى الملاحظات الهامة فى دراسة التوزيعات هى أن الكلمات ذات التكرار المنخفض لا تتبع هذا النموذج السابق أى أن ما يسمى بالرقم الثابت سيتغير ويكون أكثر اختلافاً، ومن هذه النقطة قام الباحث بوث (Booth, A. D., 1967...Cited by Cleveland, 1990) بعدة اختبارات، اقترح بعدها تعديلاً لقانون زيف الثانى كما يلى:

$I_1/I_n = (n + 1) / 2$

حيث يعتبر I_1 العدد الكلى للكلمات ذات التكرار للرقم (١)، I_n هى العدد الكلى للكلمات ذات التكرار (n) وواضح من العرض السابق أن القانونين السابقين لزيف مختلفان، نظراً لأنهما يتضمنان نموذجين سلوكيين مختلفين، ومن هنا قام الباحث جوفمان Gaffman بوضع افتراض يتضمن أنن هناك منطقة حرجة حيث يتحول عندها السلوك من الكلمات عالية التكرار والتى تتفق مع قانون زيف Zipf إلى الكلمات منخفضة التكرار كما تنبأ بها العالم بوث Booth فى التعديل الثانى أعلاه وافترض جوفمان أن هذه المنطقة يمكن أن تحتوى على كلمة واحدة أو عدداً من المصطلحات، والتى تتميز بطبيعة تحميل محتوى عالٍ وذلك فى علاقتها بالنص كله، وقد بنى جوفمان افتراضه هذا على حقيقة مؤداها أن الكلمات الوظيفية تحتل دائماً المراكز الأولى لقائمة تكرار الكلمات وهذا جزء من التركيب اللغوى كما يراه جوفمان، أما الكلمات منخفضة التكرار فتعكس الأسلوب وثراء مصطلحات المؤلفين كأفراد، وبالتالى فإن المرحلة الانتقالية لا تعكس متطلبات اللغة الأساسية ولا تعكس الأسلوب المتغير، ولكنها تمثل التجسيد الدلالى للوثيقة. Semantic embodiment of the document.

ومن الممكن فى رأى جوفمان أن النقطة الحرجة للتحول تقع حيث يقترب عدد مختلف الكلمات إلى الرقم (١) أى أن معادلة بوث Booth التالية:

$$I_1/I_n = n\,(n+1)/2$$

ستصبح $\qquad I_1/I_1 = n\,(n+1)/2$

حيث تقترب In من الرقم (١)

والحل لقيمة (n) من المعادلة التربيعية يكون كما يلى:

$$n = \frac{-1+\sqrt{1+8I_1}}{2}$$

وبالتالى فإن المنطقة الحرجة يمكن تحديدها عند النقطة التى يكون فيها حدوث تكرار للكلمات هو (n) وخلاصة ما سبق أن قوانين زيف Zipf تتنبأ بسلوك الكلمات وأن أكثر التطبيقات المباشرة يمكن أن تكون فى مجال التكشيف

الآلى، ويمكن فيما يلى توضيح بعض المحاولات الأساسية لاستخدام الحاسبات فى تكشيف الوثائق.

فالتكشيف الآلى هو تحديد مصطلحات كشفية بناء على أساس كمى، حيث يتم تلقيم العنوان أو المستخلصات أو النصوص فى الحاسب ويتم توليد المصطلحات الكشفية بطريقة آلية. وإذا ما قمنا بدمج قانون زيف Zipf بالنقطة الحرجة لدى جوفمان Goffman فسيعتبر ذلك أحد الأمثلة العملية التى يمكن أن تتم وذلك طبقاً للخطوات التالية :

(أ) قم بتلقيم الوثيقة.

(ب) سيقوم الحاسب بعدّ الكلمات ووضع رتبة لها طبقاً للقانون الأولى لزيف Zipf

(ج) يتم حساب النقطة الحرجة التى حددها جوفمان Gaffman.

(د) قائمة استبعاد الكلمات يتم استخدامها لاستبعاد كلمات مثل The/ and/ but/et ونظرا لأن المنطقة الحرجة تحتوى على مصطلحات الوثيقة التى تحمل المحتوى المعلوماتى أكثر من غيرها فإن هذه المنطقة تدلنا على مصطلحات التكشيف الآلى (Cleveland, D., 1990, P. 226) ولقد تمت عبر السنين دراسات وتغييرات عديدة لتحسين أسلوب تكرار الكلمات، وعلى سبيل المثال فقد اختزن المكنز فى الحاسب الآلى، ثم يتم مضاهاه الكلمات الأساسية المستمدة من الأساليب التكرارية على المكنز بطريقة آلية وذلك لإضافة بعض المترادفات والمصطلحات المتعلقة.

وهناك أسلوب آخر يتعلق بوزن المصطلحات، أى زيادة أهميتها النسبية وذلك بناء على وضعها فى نص الوثيقة، وعلى سبيل المثال فإن الكلمات التى توجد فى الجزء الخاص بالنتائج يمكن أن يكون لها أهمية أعلى بالنسبة للقارىء.

وهناك أيضاً مدخل استخدام مقاييس التقارب Association measures فالأساليب الإحصائية يمكن أن تستخدم لاكتشاف كيفية ظهور الكلمات جنباً إلى جنب أى فى تزواج In Pairs سواء تالية لبعضها أو تواجدها فى نفس الجملة، هذه الكلمات يمكن أن توضع فى الكشاف الناتج. وخلاصة الاستعراض السابق أن أساليب تكرار الكلمات تستخدم الحاسب الآلى لعدّ الكلمات فى النص ثم تقوم باختيار الكلمات الدالة على أساس درجة تكرارها، وهذه الكلمات الدالة تصبح الواصفات

The Descriptors وهناك مداخل أخرى كذلك على المستوى الدلالى Semantic وعلى المستوى التركيبى Syntactical ، فعلى المستوى الدلالى هناك اتجاهان يستخدمان بكثرة أولهما استخدام القاموس أو المكنز المختزن، وهنا يقوم الحاسب بتصغير الكلمات لجذورها Stems ثم العمل من خلال المكنز لبناء مصطلحات كشفية صحيحة ويلاحظ هنا أن نوعية المكنز هو أساس نجاح العملية، أما الاتجاه الثانى فيتعلق بالمحاولات الرامية لإنشاء قطاعات تصنيفية Classifiation groups بطريقة آلية من النص ثم استخدام هذه الكلمات القريبة من بعضها فى عملية التكشيف أما على المستوى التركيبى Syntactical فهناك محاولات لتطبيق الأفكار والأساليب التى طورها علماء اللغة Linguists ولكن عالم التكشيف لم يطبق كثيراً من هذه الأساليب فى حل مشكلاته.

ثالثاً – التكشيف بالاستعانة بالحاسب الآلى:

معظم نظم التكشيف الآلى ليست فى الحقيقة آلية، بمعنى استبدال الحاسبات بالإنسان، ولكن المقصود منها معاونة المكشف الإنسان، ولعل المصطلح الأفضل الذى يعبر عن هذا المفهوم: هو أنها نظم تستعين بالآلة، وهناك بصفة عامة مدخلان رئيسيان لهذا النوع من التكشيف الذى يستعين بالآلة وهما:

(١) يستخدم الحاسب الآلى لتقديم أنواع من العرض على الخط المباشر الذى يساعد فى الإسراع من عمليات المكشف، كما يمكن عن طريق الآلة اكتشاف بعض الأخطاء التى يقوم بها المكشف كاستخدام مصطلحات غير معيارية، وبالتالى يُحسِّن المكشف الإنسان من عمله.

(٢) تستخدم برامج الحاسبات الآلية لقراءة النص (رّبما قراءة العناوين أو المستخلصات فقط)، ثم اختيار مصطلحات كشفية عن طريق الاقتباس أو التعيين، والمصطلحات التى يتم تعيينها أو اختيارها بهذه الطريقة يمكن أن تراجع بواسطة المكشف الإنسان الذى يمكن أن يضيف نقاط إتاحة أخرى لا يستطيع البرنامج أن يعينها، فضلاً عن حذف أو إلغاء بعض المصطلحات التى تعينها هذه البرامج.

وعلى هذا فيمكن التمييز بين عملية التكشيف بالاستعانة بالحاسب الآلى، والتكشيف الآلى، وذلك على اعتبار أن الحاسب الآلى فى الحالة الأولى يقوم

بالأعمال الروتينية بينما يستمر الإنسان فى القيام بالعمل الفكرى للتكشيف (Cleveland. 1990. p. 228) ولقد استخدمت الحاسبات بنجاح منذ زمن غير قصير، وذلك لمعاونة التكشيف الإنسانى اليدوى وتجهيزه وإنتاج القوائم للمطبوعات الكشفية، وهنا ثلاث خطوات تتم عادة بالنسبة للتكشيف الذى يستعين بالحاسب وهى:

(١) يقوم الإنسان المكشف بفرز الوثائق اللازمة للتكشيف بناء على سياسة موضوعة، ثم اختيار الأجزاء الخاصة بالتكشيف (كالعنوان أو رؤوس الموضوعات الموضوعة لكل جزء فى الوثيقة أو الجمل التى تغطى هدف معين أو المناهج المستخدمة أو النتائج التى تم التوصل إليها، وقد يتم اختيار الجمل الأولى أو الأخيرة فى الفقرات وغير ذلك) ويلاحظ هنا إمكانية قيام موظف غير مهنى بهذا العمل، وذلك بعد تدريبه لدرجة مناسبة من المهارة.

(٢) إدخال هذه المواد فى الحاسب، ثم تشغيل الآلة لاستخدام أحد أساليب التكشيف الآلية المعيارية (كتلك المعتمدة على قانون زيف Zipf) وذلك لإنتاج المصطلحات الكشفية.

(٣) القيام بعملية تحرير للنتائج مع عمل التغييرات أو التعديلات الضرورية وهذه العملية هى مهمة المكشف المهنى الخبير.

وكما يقول كليفيليند (Cleveland, D., 1990, p. 229) فإن هذا الاتجاه يعبر عن التوازن الاقتصادى بين قوة العمل غير الماهرة والماهرة والحاسب الآلى، حيث يؤدى العمل الفكرى بواسطة المكشف المهنى بينما تتم التطبيقات الصعبة اللازمة لإتمام عملية التجهيز بواسطة الحاسب.

هذا وبرامج التكشيف للحاسبات المصغرة بدأت تغزو الأسواق منذ بداية الثمانينيات مع استمرار تحسين هذه البرامج. هذا وقد قامت لندا فيترز (Fetters, L. K., 1987) بوضع إطار لقواعد مرشدة لتقييم برامج التكشيف كما يلى:

(١) التشكيل **Formatting**

أ – القيام بالتشكيل الآلى للكشاف النهائى بأسلوب معترف به.

ب– إنشاء عدد مقبول من المداخل الفرعية (٣–٧) بطريقة آلية.

ج – استبعاد المداخل الرئيسية المكررة.

د – التجميع الآلى لمراجع الصفحات بالنسبة للمداخل المتشابهة.

(٢) إنشاء وتحرير المداخل

أ – كفالة طول مناسب لكل مدخل.

ب– الاستدعاء اليسير للبيانات التى سبق إدخالها ثم القيام بتحريرها على الشاشة.

ج – كفالة وسيلة لعرض وطبع المداخل عند أى نقطة فى عملية التكشيف.

د – كفالة وسيلة لفرز ونسخ رؤوس الموضوعات أو الرؤوس الفرعية السابق استعمالها.

هـ – إمكانية اختزان الكشاف النهائى كملف كلمات مجهزة، على أن يتم الاحتفاظ بالتسجيلات الأصلية للاستخدام المستقبلى.

(٣) الفرز:

أ – نظام الفرز يعالج الأحرف الطباعية الصغيرة والكبيرة بطريقة متساوية (فى اللغة الإنجليزية).

ب– طريقة لوضع علامات للرموز أو الكلمات التى لا تعتبر جزءاً من نظام الفرز كالبنود والفقرات وحروف الجر.

ج – القدرة على الفرز حرف بحرف أو كلمة بكلمة.

د – القدرة على الفرز بواسطة المدخل الرئيسى والمستويات الأخرى من المداخل الفرعية.

(٤) مستلزمات الطباعة :

أ – إمكانيات رسم الخطوط تحت الكلمات أو التنفيذ بحروف مطبعية سوداء أو الرموز أو الأرقام المكتوبة تحت أو فوق الكلمات.

ب – إمكانيات تغيير أو إدخال الأكواد الخاصة بالطباعة كما يحتاجها كل ناشر.

(٥) تركيم أو دمج الكشافات:

أ – إمكانية معالجة مداخل كافية للمشروعات الكبيرة أو متعددة المجلدات.

ب- إمكانية تركيم أو دمج الكشافات المفردة فى كشاف واحد كبير وهناك برامج معدة لمواجهة هذه المتطلبات الكشفية (Cleveland, D., 1990, P. 230).

رابعاً: التكشيف الاقتباسى الآلى والتكشيف التعيينى الآلى:

هناك نوعان من التكشيف هما التكشيف بالتعيين Assigned Indexing والتكشيف الاشتقاقى أو الاقتباسى Derived or Extraction Indexing حيث يعنى التكشيف بالتعيين قيام المكشف باختيار الواصفات لتمثيل المفاهيم الموجودة بالوثيقة بينما يتضمن التكشيف الاشتقاقى أو الاقتباسى استخدام المكشف للكلمات الفعلية للمؤلف على اعتبار أنها واصفات Descriptors وذلك دون تعديل أو تحوير من المكشف، والتكشيف الآلى يعتمد على افتراض مؤداه أن كلمات النص وعلاقتها بعضها ببعض تعتبر كافية لتمثيل مفاهيم المحتوى وهذا تكشيف اشتقاقى.

(١) التكشيف الاقتباسى الآلى Automatic Extraction Indexing :

يحاول المكشف الإنسان اختيار بعض التعبيرات النصية التى تبدو كمؤشرات جيدة لما تهدفه وتقصده الوثيقة. ويفترض فى هذه الكلمات المعبرة أن عدد مرات تكرارها ستكون أكثر، كما أن موقع هذه الكلمات له أهمية (فى العنوان، فى الملخص، فى شروحات الأشكال)، فضلاً عن الكلمات داخل السياق Context .

وإذا ما تم وضع النص فى شكل مقروء آلياً، فمن الواضح أن الحاسب الآلى يمكن أن يُبرمج للقيام بالتكشيف عن طريق الاقتباس، مستخدماً نفس المعايير السابقة المتصلة بدرجة تكرار الكلمات وكيفية وضعها فى النص أو دلالتها فى المتن، فهناك بعض البرامج البسيطة التى يمكن كتابتها لعدّ درجة حدوث أو ورود الكلمات فى النص لا سيما عند وضع ما يسمى بقائمة الاستبعاد (stop list) من هذا النص لاستبعاد الكلمات غير ذات الدلالة (أدوات الربط).

ثم يتم بعد ذلك وضع هذه الكلمات فى ترتيب معين طبقاً لدرجة تكرار هذه الكلمات، حيث يتم اختيار الكلمات التى فى قمة القائمة لاختيارها كمصطلحات تكشيفية للوثيقة وهناك بعض البرامج الأكثر تعقيداً وهى التى يمكنها اقتباس جمل، والتى تحدث بطريقة ذات دلالة فى النص، وبالتالى فيمكن تمثيل الوثيقة بتوليفات من الكلمات والجمل.

وهناك اتجاه آخر لا يركز على اختيار الكلمات أو الجمل، ولكن البرامج فى هذه الحالة تكتب لاختيار جذور الكلمات فكلمة Heat (حرارة) يمكن اعتبارها جذر يتم اختياره والاحتفاظ به بدلاً من متغيرات هذه الكلمة مثل Heated, Heating، والبرامج الآلية هذه تقوم فقط باستبعاد نهايات الكلمات (باللغة الإنجليزية مثل ing, ed)، ومن الطبيعى أن الكلمات أو الجمل أو الجذور يمكن أن تعطى جميعها أوزان تعكس مدى تكرارها فى الوثيقة (أى أن جذور كلمة Heat السابقة يمكن أن يكون لها وزن عددى مرتبط بحقيقة ظهورها فى النص إثنتى عشر مرة) (Lancaster, F., 1991, p. 221).

هذا ويمكن استكمال معايير التكرار Frequency بمعايير أخرى، فعلى سبيل المثال اقترحت الباحثة باكسندال (Baxendale, 1958) أن الجملتين الأولى والأخيرة من كل فقرة يمكن أن يتم إدخالهما فى المعالجة، لأن إحدى دراساتها قد أظهرت أن الكلمة الأولى هى المعبّرة عن الموضوع فى حوالى ٨٥% من الأحيان، أما الجملة الأخيرة فقد عبّرت عن الموضوع فى حوالى ٧% من الحالات، وجملة الموضوع هى الجملة التى تقدم معظم المعلومات المتصلة بالمحتوى، هذا بالإضافة إلى أنه فى الفترات الزمنية الأولى للتكشيف الآلى، اتبعت عدة طرق مختلفة لتحديد أجزاء النص ذات الثراء بالمعلومات فقد تبحث برامج الحاسب الآلى عن العناصر التابعة للكلمات المفتاحية كالنتائج والملخص، وأجزاء النص التى تشمل أكثر الأسماء حدوثاً. ولعلنا نلحظ بعض العيوب فى استخدام الكلمة الواحدة أو عدد تكرار الفقرات فى اختيار المصطلحات، ذلك لأن هذه المصطلحات التى توجد بكثرة فى الوثيقة قد لا تعتبر المصطلحات التى تميز هذه الوثيقة عن غيرها فى قاعدة البيانات.

ولنأخذ مثالاً للتدليل على ذلك: فكلمتى المكتبات والمعلومات سوف لا تكون كلمات مميزة داخل النص الخاص بعلم المكتبات والمعلومات، وعلى ذلك فإن كلمة مكتبة أو مكتبات قد توجد ١٢ مرة بينما توجد كلمة مثل Asbestos (٤) مرات فقط، ومع ذلك فإن الكلمة الأخيرة تعتبر كلمة مميزة لأنها نادراً ما تحدث فى الإنتاج الفكرى لعلم المكتبات، ونفس هذا المصطلح Asbestos المادة العازلة ستكون ذات دلالة عالية فى مجموعة عن هذا الموضوع، حتى إذا حدثت لمرة واحدة فى الوثيقة.

هذا وتكرار الكلمات المعينة فى وثيقة ما لا يعتبر التكرار الوحيد الذى يهتم به التجهيز المحسب للنص، ذلك لأن تكرار الكلمة فى قاعدة البيانات كلها قد يكون له دلالة أكبر، كما هو الحال بالنسبة لكلمة Asbestos فى الإنتاج الفكرى لعلم المكتبات، أى أنها قد تعنى المكتبة فى قاعدة البيانات لشركة Asbestos، وقد يكون من غير الضرورى حساب تكرار كلمة معينة فى قاعدة البيانات الكاملة للنصوص، ولكن حساب التكرار الذى تحدث به الكلمة فى الفهرس المقلوب Inverted file قد يكون أهم [أى عدد تكرارت كلمة فى علاقتها بعدد تكرارات جميع الكلمات فى الفهرس أو الملف].

وهناك من يتعامل مع معيار التكرار ليس بالطريقة المطلقة، ولكن بالطريقة النسبية، فعلى سبيل المثال إذا حدث تكرار لكلمة معينة لمرة واحدة فى عدد خمسة آلاف كلمة من مقال الدورية فهذا التكرار المفرد يعتبر ذا دلالة أكبر إذا حدثت نفس الكلمة خمس مرات فقط فى قاعدة بيانات تشمل عشرة ملايين كلمة وعلى ذلك نجد أن كل هذه الأمثلة تعكس صعوبة التكشيف الآلى المبنى على عدد تكرارات الكلمات، فضلاً عن أن الاسترجاع الفعال للمعلومات يتطلب تجمعات للكلمات، وليس مجرد كلمة واحدة (أى مضاف ومضاف إليه، صفة وموصوف، عطف ومعطوف كالخير والشر ... إلخ) .

هذا ويمكن القيام باقتباس مصطلحات من النص عندما يتم مضاهاة هذه المصطلحات بمصطلحات مقبولة فى قاموس مختزن، وقد كانت هذه الطريقة أساس التكشيف الآلى فى السبعينيات فى مركز توثيق الدفاع الأمريكى، وهذه الطريقة فى محورها تعكس خيوط كلمات Word strings توجد فى العناوين والمستخلصات، ويتم مضاهاتها بقاعدة معلومات اللغة الطبيعية Natural Language Data Base (NLDB) وقد قام بعض الباحثين بمقارنة نتائج التكشيف الذى يستعين بالآلات بنتائج التكشيف الإنسانى، وقد انتهى هؤلاء الباحثون إلى أن التكشيف الذى يستعين بالآلة دون عملية التحرير Unediting يمكن أن تؤدى إلى مستويات استدعاء تقترب من الاستدعاء الذى يحدث فى التكشيف الإنسانى، وأن الدقة التى تتحقق مع التكشيف الذى يستعين بالآلات قد تصل فى بعض الأحيان إلى التكشيف الإنسانى، ولكن التكشيف الآلى الذى يتعرض لعملية التحرير يحقق نتائج

استدعاء يمكن اقترابها من التكشيف الإنسانى، ولكن فى الوقت ذاته يحقق الدقة بطريقة أفضل من التكشيف الإنسانى، ولعل هذه المقارنات السابقة تعكس مقارنة التكشيف باللغة الطبيعية فى مقارنتها بالتكشيف بالمصطلحات المحكومة.

(٢) التكشيف الآلى بالتعين Automatic Assignment Indexing :

يعتبر اقتباس الكلمات أو الفقرات من الوثائق مهمة يمكن أن يحققها الحاسب الآلى بنجاح، وهناك ميزة واضحة للاقتباس الآلى تتفوق بها عن الاقتباس الإنسانى، وهى عملية الانتظام (Consistent) ومع ذلك فمعظم التكشيف الإنسانى ليس تكشيفاً اقتباسياً، ولكنه تكشيف بالتعيين، والقيام بهذه المهمة بواسطة الحاسبات يعتبر أكثر صعوبة، والطريق الواضح للقيام بالتكشيف بواسطة التعيين والاستعانة فى ذلك بالحاسب هى إيجاد مجموعة من الكلمات أو الفقرات التى تمثل سمات Profile لكل مصطلح يتم تعيينه أو وضعه، والكلمات والفقرات التى توجد ضمن ما يسمى بالسمات هى الكلمات أو الفقرات التى تميل للحدوث بكثرة فى الوثائق التى سيقوم المكشف الإنسانى بتعيين هذا المصطلح فيها، والسمات مثلاً بالنسبة لمصطلح المطر الحامض Acid Rain قد تشمل فقرات مثل ترسيب الحامض Acid Precipitation أو تلوث الهواء Air Pollution أو ثانى أكسيد الكبريت Sulfur dioxide وغيرها – (Lancaster, F., 1991, p. 225) وهناك مشكلات عديدة فى بناء هذه السمات حول مصطلحات معينة، من أجل ذلك فإن المحاولات الأولى لتعيين أو وضع مصطلحات بطريقة آلية كانت غير ناجحة، حتى مع عدد قليل جداً من المصطلحات (Borko & Bernick 63) ومع ذلك فقد شهدت السنوات الخمس والعشرون السابقة تطورات ناجحة وأصبح من الممكن فى الوقت الحاضر القيام بالتكشيف بالتعيين بنجاح أكبر. ولعل برنامج Biosis يعتبر واحداً من البرامج المعقدة للتكشيف الآلى بالتعيين، وقد ناقشه كل من خلودتس وستوكولوف.

(Vleduts - Stokolov, 1987,... Cited by Lancaster, F., 1991, p. 226)

فقد قام بمضاهاة الكلمات التى تظهر فى عناوين مقالات الدوريات مع مصطلحات دلالية Semantic Vocabulary تتكون من حوالى (١٥) ألف مصطلح بيولوجى، وهذه بدورها قد تم وضعها فى خريطة لرؤوس المفاهيم Concept Heading

تتكون من ٦٠٠ مفهوم (أى أن هذه الخريطة تحتوى على رؤوس موضوعات عريضة نسبياً)، ونتيجة لذلك أصبح من الممكن تعيين هذه الرؤوس المفاهيم بواسطة الحاسب الآلى على أساس الكلمات والفقرات التى تحدث فى العناوين، وقد أشار الباحث فيلودوتس فى المرجع السابق إلى أن حوالى ٦١% من الرؤوس المفاهيم والتى تم تعيينها بواسطة الإنسان يمكن أيضاً تعيينها بواسطة الحاسب الآلى، وذلك اعتماداً على العناوين وحدها.

ومع ذلك فيجب أن نشير إلى أن البرامج الحالية لا تحقق مستوى أداء عالٍ ذلك لأنها ناجحة فى حوالى ٨٠-٩٠% بالنسبة للتعيينات الأولية والثانوية (ومعنى ذلك إمكانية تعيين من ٨٠-٩٠% من نسبة ٧٥% التى يمكن تعيينها نظرياً اعتماداً على العناوين، والمستوى الأولى والثانوى هو خطة – لوزن مستوى المصطلحات، حيث يحدد نظام بيوسيس Biosis ثلاثة مستويات فى هذا الوزن الأولى والثانوى ومن الدرجة الثالثة، ومعنى ذلك أيضاً أن هذه البرامج تفشل فى تعيين بعض المصطلحات التى يجب تعيينها، والتى أمكن تعيينها بواسطة الإنسان، والعكس صحيح أيضاً فهناك بعض التعيينات الزائدة التى لا لزوم لها.

وهناك اتجاه آخر فى هذا المجال قام به الباحث تروبكين (Trubkin, 1979... Cited by Lancaster, F., 1991, p. 227) للتكشيف الآلى للمستخلصات فى قاعدة معلومات التجارة والأعمال ABI/Infrom وذلك خلال الفترة من ٧١-١٩٧٧م. وقد استخدم هذا الباحث ما يسمى بمصطلحات الجسر Bridge vocabulary والتى تصل إلى حوالى ١٩ ألف مصطلح، وهذه تساعد فى التحويل من التعبيرات الموجودة فى النص إلى المصطلحات المحكومة، وفى هذه الحالة فإن وجود المصطلح فى العنوان أو المستخلص ولو لمرة واحدة يكفى لتعيين مصطلح محكوم، وقد نتج عن ذلك أن إجراءات التكشيف الآلية تميل إلى تعيين مصطلحات أكثر للمادة مما يقوم به عادة الإنسان المكشف (متوسط ١٦ مصطلح لكل مادة بالمقارنة بـ٨-١٢ للإنسان).

وهناك عمل مشابه لبيوسيس Biosis ، وهى إجراءات التكشيف الآلى التى قام بها معهد البترول الأمريكى وقد كان الهدف هو وضع طرق تسمح للحاسب الآلى بتعيين المصطلحات المحكومة من مكنز API على أساس نص المستخلصات.

ويشير الباحث برينير Brenner وزملاؤه إلى أن الصيغة الأولى للنظام تم فيها تعيين ٤٠% فقط من المصطلحات التى كان يعينها المكشف الإنسان، بالإضافة إلى تعيين عدد من المصطلحات غير المرغوبة (cited by Lancaster, 1991, p. 226) ولكن الباحثين تعلموا من هذه التجربة التفاؤل بإمكانية قيام الإجراءات الآلية بتعيين حوالى ٨٠% من المصطلحات التى يجب تعيينها، وأن هذه الإجراءات المطورة ستساعد على تخفيض له دلالته بالنسبة للتعيينات غير المرغوب فيها. وقد أشار إلى ذلك الباحث ميوتينز وزملاؤه ..Mutinz, 1987) (cited by Lancaster, 1991, p. 227 حيث قاموا بوصف هذه التطورات فضلاً عن توضيح المشكلات التى تتصل بوضع الخريطة المفهومية من التعبيرات النصية إلى مصطلحات المكانز. وخلاصة ما تم استعراضه سابقاً أنه على الرغم من التحسن الملحوظ فى التكشيف الآلى بالتعيين خلال السنوات الخمس والعشرين الماضية، فلم نصل بعد إلى نقطة تمكننا من تعيين المصطلحات فيها من بين عدد كبير من المصطلحات (مثلاً اختيار عشرة آلاف من الواصفات فى أحد المكانز) دون الاستعانة بالتدخل الإنسانى.

وفى واقع الأمر فالتكشيف الآلى بالتعيين لم يُعد له اهتمام كبير بين المتخصصين إلا بالنسبة لإنتاج الكشافات المطبوعة، خاصة وقد كانت بدايات هذا النوع من التكشيف مكلفة للغاية بالنسبة لاختزان وتجهيز الكميات الضخمة من النصوص بواسطة الحاسب الآلى، وبالتالى فإن أى طريقة كانت تصغر النص، وتجعله أكثر اختصاراً كان لها ما يبررها.

ولعل كشاف الكتاب العادى، يعتبر شكلاً متخصصاً من الكشاف المطبوع ويعود هذا العمل منذ حوالى ٢٥ عاماً إلى الباحثة ارتاندى ... Artandi, 1963) (cited by Lancaster, F., 1991, p. 228 التى أنتجت كشافات الكتب بواسطة الحاسب الآلى فى مجال الكيمياء، فلكل مدخل كشفى (مصطلح تم التعبير عنه) قامت Artandi باستخلاص قائمة من الفقرات المرتبطة بحيث إذا ما وجدت أى واحدة منها فى صفحة النص، فإن ذلك سيؤدى إلى اختيار واحدة من المداخل الكشفية لهذه الصفحة، وقد زعمت أرتاندى Artandi أن الكشاف المنتج بهذه الطريقة يمكن أن يكون على مستوى مماثل فى نوعيته بالكشاف المعد بواسطة

الإنسان، وإن كان الكشاف الآلى أكثر تكلفة من غير شك، لأن ذلك يعود إلى وضع النص فى شكل مقروء آلياً.

ومعنى ذلك أن عوامل التكلفة فى الوقت الحاضرلم تعد تؤيد العمل الفكرى الإنسانى، نظراً لأن كل الطباعة تتم فى الوقت الحاضر من مدخلات مقروءة آلياً، ومع ذلك فما زالت مشكلة إنتاج كشافات الكتب آلياً أمراً عسيراً وأكثر صعوبة من بدايات ارتاندى Artandi ذلك لأنه فى أى مجال موضوعى محدود، هناك مصطلحات تعبير كثيرة جداً نحتاجها، ولكل واحدة منها بعض المصطلحات الخرائطية الكبيرة، فضلاً عن ضرورة تحديث المصطلحات هذه وتلك حتى تعكس التطورات الجديدة والمصطلحات المتغيرة فى المجال.

خامساً – تجارب فى التكشيف الآلى للوثائق العربية فى نظام استرجاع المعلومات:

تشير الباحثة فى هذا الجزء إلى بعض التجارب المنشورة فى مجلة الجمعية الأمريكية لعلم المعلومات JASIS الأولى عام ١٩٩٤ (AL- Kharashi, I., 1994) والثانية عام ١٩٩٧م (Hmeidi, I., 1997) وقد قام بالتجربة الأولى الخراشى بمدينة الملك عبد العزيز للعلوم والتقنية بالرياض (الإدارة العامة لخدمات المعلومات) بالاشتراك مع مارثا إيفنز من قسم علوم الحاسب بمعهد إلينوى للتكنولوجيا بشيكاغو، أما التجربة الثانية فقد قام بها الحميدى بالمعهد الأردنى للعلوم والتقنية فى أربيد بالأردن بالإشتراك مع كل من غسان كنعان ومارثا إيفنز أيضاً.

وفى التجربة الأولى تم تصميم نظام تجريبى يستخدم فيه الحاسب الشخصى فى استرجاع المعلومات العربية، وذلك للتعرف على عمليات التكشيف والاسترجاع بالنسبة للبيانات الببليوجرافية العربية، وقد تمت سلسلة من التجارب (٢٩) سؤالاً لقاعدة البيانات المكونة من (٣٥٥) تسجيلة ببليوجرافية وهذه القاعدة تغطى مجال علم الحاسب والمعلومات من بنك المعلومات الموجود بمدينة الملك عبد العزيز للعلوم والتقنية بالرياض، وقد كشفت هذه التجارب أن استخدام الجذور roots والجذوع Stems كمصطلحات كشفية، يقدم لنا نتائج استرجاعية أفضل من مجرد استخدام الكلمات ونتائج الجذور تكون فى أدائها أفضل من الجذوع. عند مستويات

الاستدعاء المنخفض، ولكن الجذور تكون أفضل بالتأكيد عند مستويات الاستدعاء العالى، كما تم استخدام عدد من المعاملات (The Cosine, Dice and Jaccard Coefficients) وقد أظهرت المعاملات الثلاث نفس ترتيب الوثائق المسترجعة بالنسبة لكل سؤال وبالتالى فإن استخدام أى معامل واحد بسيط يكفى وقد أجريت التجارب على حاسب مصغر متلائم (IBM/AT) أما نظام استرجاع المعلومات العربى Micro Airs فقد تمت كتابته بلغة تيربو/س TURBO/C, Version 2.5 ويلاحظ فى هذه التجربة تناول المشكلات التالية:

(أ) الوصول إلى النص العربى/ الإنجليزى ومعالجته وعرضه.
(ب) تكشيف وفرز المصطلحات العربية.
(ج) تكشيف واسترجاع البيانات العربية باستخدام أشكال مختلفة من المصطلحات الكشفية وهى الكلمات والجذوع والجذور.
(د) ترتيب الوثائق باستخدام معاملات ثنائية مختلفة.

ويذهب الباحثان إلى أن النظم التشغيلية Operational Systems تحتاج إلى استبدال قاموس الكلمات ــ الجذوع/ الجذور بخوارزمية صرفية Morphology algorithm يتوفر بها الجذوع والجذوع. كما تبين أن فشل الاسترجاع يرجع إلى استخدام الكلمات الغريبة (كالمترادفات)... ويعتقد الباحثان أن استخدام مكنز متفاعل Interactive thesaurus سيؤدى إلى استرجاع وثائق صالحة أكثر، كما يقترح الباحثان لتحسين أداء نظام استرجاء المعلومات استخدام عملية ترتيب موزونة Weighted ranking process. وذلك بالنسبة لقاعدة معلومات تتوفر بها المستخلصات وليس مجرد العناوين التى استخدمت فى التجربة، كما قد يستخدم النص الكامل على الخط المباشر وهذه هى المشروعات الحالية بمدينة الملك عبد العزيز للعلوم والتقنية.

أما التجربة الثانية وهى الأحدث فقد أفادت من الدراسات السابقة والجزئية الهامة فى الدراستين وجود شخص مشترك هو مارثا إيفانز[*] وحيث توجد

[*] تشير الباحثة إلى أن تجارب التكشيف الآلى فى اللغات الأخرى تستعين بالدراسات الخاصة بالتكشيف الآلى لاسترجاع الوثائق العربية وعلى سبيل المثال فتجارب التكشيف الآلى فى نظام استرجاع المعلومات الصينى=

المراسلات لمارثا إيفانز بقسم علوم الحاسب بمعهد الينوى للتكنولوجيا بشيكاغو من أى مكان من العالم وتتكون مجموعة التجربة هنا من عدد (٢٤٢) مستخلصاً لوثائق عربية من وقائع المؤتمرات الوطنية بالمملكة العربية السعودية كمصدر، وهذه المستخلصات جميعها فى مجال نظم المعلومات وعلم الحاسب، وقام الباحثون ببناء نظام استرجاع آلى للمعلومات لمعالجة البيانات العربية، وتم تطبيق النظام بلغة س (C language) وباستخدام مترجم (GCC Compiler) ويتم تشغيله على حاسب شخصى (IBM/ PC) وقد تم فى هذه التجارب القيام بأساليب التكشيف اليدوى والآلى، واستخدمت إجراءات الاستدعاء والدقة Recall & Precision حيث أظهرت هذه المقاييس أن التكشيف الآلى يؤدى إلى نتائج شبيهة بالتكشيف اليدوى إن لم يكن أفضل منه فى بعض الحالات، ونظراً لأن التكشيف الآلى أرخص وأسرع فإن نتائج التجارب تبشر بإمكانية تحقيق تغطية أوسع من الإنتاج الفكرى بتكاليف أقل ونتائج أفضل من التكشيف اليدوى.

كما قام الباحثون كذلك بمقارنة نتائج الاسترجاع باستخدام الكلمات كمصطلحات كشفية وباستخدام الجذوع والجذور (Stems and Roots) وقد أيدت التجارب هذه النتائج التى سبق أن وصل الخراشى (Al - Kharashi) وأبو سالم Abo- Salem إليها من قبل فى هذا الخصوص ولكن بعينة أصغر من تلك المستخدمة فى التجارب هنا، وتؤيد هذه التجارب كالتجارب السابقة أن تكشيف الجذور أكثر فاعليه من تكشيف الكلمات. وأخيراً فقد أيّد الطيار[1] (١٩٩٨، ص ٢١) هذه النتائج فى دراسته الأحدث، ويرى الطيار أن أداء الدقة Precision بالنسبة للبحث بمستوى الجذر، يمكن تحسينه مع التعمق فى التحليل اللغوى للجذور العربية.

=(Tian- Long Wan, 1997, 1086) قد اشتركت فيها مارثا إيفانز بقسم علوم الحاسب بمعهد إلينوى للتكنولوجيا، كما جاءت أسماء باحثين عرب إلى جانب الخراشى مثل الباحث أبو سالم (Abu- Salem, H., 1992) والباحث الفداغى (AL- Fedaghi, S., 1989) أى أن هناك سمات وجوانب مشتركة بالنسبة للتكشيف الآلى فى اللغات المختلفة خاصة الإنجليزية والعربية والصينية.

(١) مساعد بن صالح الطيار (١٩٩٨) كفاءة التحليل الصرفى فى استرجاع النصوص العربية. مجلة مكتبة الملك فهد الوطنية، مج٤، ع١، ص ٧ -٢٣.

سادساً: أشكال واتجاهات أخرى لتحسين التكشيف الآلى:

(١) توليد الكشافات المطلوبة المضغوطة:

فى بحث لموفات (Moffat, A., 1995) فى مجال التكشيف الآلى، شرح لنا تجربته فى اختزان الكشافات المقلوبة حيث وضع لكل مصطلح يظهر فى مجموعة الوثائق، قائمة من أرقام الوثائق التى تحتوى على هذا المصطلح، ومثل هذا الكشاف لا يمكن الاستغناء عنه عند الإجابة على الأسئلة البولينية أو المرتبة بطريقة رسمية، ومع ذلك فإن بناء هذا الكشاف يعتبر عملاً غير عادى، فالطرق البسيطة التى تستخدم تركيبات البيانات فى الذاكرة لا يمكن استخدامها بالنسبة للمجموعات الكبيرة، لأنها تحتاج إلى كثير جداً من الوصول العشوائى للاختزان، كما أن الطرق التقليدية التى تعتمد على الاسطوانات Discs تتطلب كميات ضخمة من المساحة المؤقتة للملفات. ويشرح لنا الباحث موفات Moffat الجوريزم تكشيفى جديد مصمم لإنشاء كشافات مقلوبة مضغوطة كبيرة فى الموقع، وهى تعتمد أو تفيد من الأكواد المضغوطة البسيطة لإنشاء أعداد إيجابية صحيحة وقد استخدمت الأساليب الجديدة هذه فى تحويل مجموعة نصوص تصل إلى (٢ جيجا بايت) فى حوالى أربع سنوات باستخدام أقل من (٤٠ ميجا بايت) من مساحة الأسطوانه المؤقتة وأقل من (٢٠ ميجا بايت) من الذاكرة الرئيسية).

(٢) مداخل وزن المصطلحات فى الاسترجاع الآلى للنص:

يذهب الباحث الشهير جيرالدسالتون وزميله باكلى (Salton, G. et al, 1988) إلى أن الدليل التجريبى المجمع على مدى العشرين سنة السابقة يشير إلى أن نظم تكشيف النص المعتمدة على تعيين المصطلحات المفردة الموزونة المناسبة يؤدى إلى نتائج استرجاعية متفوقة عن تلك التى تحصل عليها بتمثيلات النص الأكثر تفصيلاً، ولكن هذه النتائج تعتمد بشكل حاسم على اختيار نظم وزن المصطلحات الفاعلة، وقد قدم سالتون وزميله باكلى دراسته مبيناً فيها ما اكتسباه من خبرة فى وزن المصطلحات آلياً وقدما نماذج تكشيف المصطلح الواحد Baseline Single-term والتى يمكن الإفادة منها للقيام بتحليل محتوى أكثر تفصيلاً للمقارنة.

(٣) تحسين التكشيف الآلى من خلال التعرف على الفقرات المتماثلة دلالياً والمختلفة تركيبياً:

تتناول هذه الورقة البحثية (Aladesulu. O.S., 1986) طرق تحسين عمليات تكشيف الوثائق آلياً، بحيث تقترب من المستوى الفكرى للتكشيف الإنسانى، وقد قام الباحث بوصف الجورزم معرفى دلالى Semantic recognition مع تطبيقه، ثم قام بتقييم هذا العمل اعتماداً على مقياس مقارن يسميه دالة الثقة Confidence function .

(٤) التواجد المصاحب والاستشهاد المصاحب والتزاوج الببليوجرافى:

Co – occurance, Co- citation and Bibliographic Coupling

العلاقة الواضحة التى يمكن استغلالها بواسطة الحاسب الآلى هى التواجد المصاحب Co-occurance ، وكلما زاد عدد تواجد مصطلحين مع بعضهما (فى نص الوثائق أو فى قائمة المصطلحات التى يتم تعيينها للوثائق) فمن المحتمل أنهما يعالجان مادة موضوعية متشابهة، ويمكننا التعبير عن ذلك بنتيجة منطقية، وهى إذا لم يحدث المصطلح (س) أبداً دون المصطلح (ص) وإذا لم يحدث المصطلح (ص) بدون المصطلح (س) (وهذه حالة نادرة للغاية) فإن المصطلحين يعتمدان أحدهما على الآخر بطريقة كاملة، بل ويمكن أن يستبدل أحدهما بالآخر فى البحث، وإلى جانب الارتباط المباشر يميل المصطلحات (ص)، (س) إلى الحدوث مع بعضهما، وبالتالى فإن الارتباط غير المباشر بين المصطلحات يمكن استخلاصه على أساس ازدواجية حدوث البيانات] لنفترض على سبيل المثال أن المصطلح [د] لا يحدث مطلقاً فى أى قاعدة بيانات دون المصطلح [و]، إلا أن المصطلح [ت] يميل لعدم الحدوث بدون المصطلح [و]، إلا أن المصطلح [د]، [ت] لا يتم حدوثهما المتزامن أبداً فى الوثائق، فقد نستنتج من ذلك أن هناك بعض العلاقات التى توجد بين مصطلح [د] ومصطلح [ت] وهما متعلقان ببعضهما نظراً لأن كلاً منهما يحدث بالتزامن وبشدة مع المصطلح [و]، وفى كل الاحتمالات يكون المصطلحان [د] ومصطلح [ت] مترادفان تماماً فى النص، والمترادفات تميل إلى عدم الحدوث مع بعضها، ومع ذلك فإن المصطلحات التى تحدث معها بالتزامن ستكون متشابهة جداً.

وفى الحقيقة فإن درجة الارتباط بين مصطلحين يجب أن لا تحسب على أساس التكرار البسيط لتزامن الحدوث، ولكن يجب أن تحسب على أساس تكرار تزامن الحدوث فى علاقته بتكرار حدوث كل مادة.

فعلى سبيل المثال إذا كان المصطلحان أ، ب يحدثان بالتزامن ٢٠ مرة فى قاعدة البيانات، بينما يوجد المصطلح (أ) عشرة آلاف مرة، والمصطلح (ب) خمسون ألف مرة، فإن عامل الترابط بين (أ)، (ب) سيكون ضعيفاً، ومن جهة أخرى إذا افترضنا أن المصطلح (أ) يحدث خمسين مرة والمصطلح (ب) يحدث ٢٥ مرة، فإن حدوثهما المتزامن قد يكون عشرين مرة، أى أن عامل الارتباط فى هذه الحالة سيكون كبيراً نظراً لأن المصطلح (ب) سوف لا يحدث عادة بدون مصطلح (أ)، ونصف حالات حدوث المصطلح (أ) تتطابق مع حدوث مصطلح (ب)، وبالتالى فإن درجة العلاقة للمصطلحين تعرف عادة بالمعادلة البسيطة

(Relatedness) R

$$R = \frac{A \text{ and } B}{A \text{ or } B} \qquad ع = \frac{أ + ب}{أ \text{ أو } ب}$$

وعندما تزيد العلاقة ع، R على عتبة (قيمة) Threshold محدودة مسبقاً فيفترض أن المصطلحين متعلقان ببعضهما.

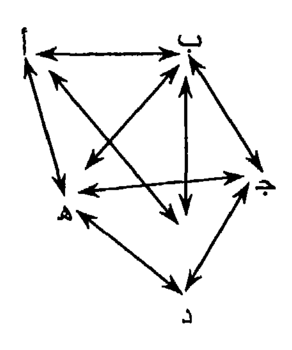
هذا وبيانات الحدوث المتزامن يمكن أن تستخدم بطريقتين داخل نظام الاسترجاع الآلى:

(أ) شبكة من الارتباطات بين المصطلحات يمكن وضعها واختزانها.

(ب) أقسام محددة من المصطلحات أى تجمعات من المصطلحات يمكن تحديدها واختزانها اعتماداً على الارتباطات المستمدة من الشبكة والحالة الأولى تحتوى على مدخلات المصطلحات بواسطة الباحث على هيئة قائمة مصطلحات أو كبيان فى الفقرة أو الجملة وبالتالى فإن هذه المدخلات من المصطلحات يمكن أن تفصل وتتوسع آلياً لإنتاج قائمة موسعة من مصطلحات البحث، أما فى التطبيق الثانى فإن أى كلمة تحدث فى بيان البحث يمكن استبدالها بمجموعة

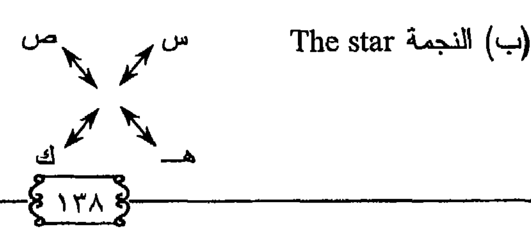
من الكلمات التى تنتمى هى إليها، وهذا الإحلال قد يكون آلياً أو تحت سيطرة الباحث نفسه، هذا والأقسام Classes التى تتكون بواسطة الإجراءات الإحصائية قد تكون أقل نقاءً من تلك الموجودة فى المكنز التقليدى، ذلك لأنها مجموعة الكلمات التى تحدث بشدة بطريقة متزامنة co-occurance أى أن هذه تشمل الجنس والنوع والجذر والكل وغيرها من العلاقات، والقضية هنا هى هل سيصلح هذا القسم وأن يكون مفيداً فى الاسترجاع أم أنه سيسبب انخفاضاً حاداً فى الدقة Precision خاصة إذا ضم القسم مصطلحات غير متجانسة، ويلاحظ فى المناقشات السابقة أنها استخدمت فقط الطرق التى تشكل فيها جماعات أو أقسام المصطلحات على أساس الوثائق التى توجد فيها. كما أن البيانات التى تتيح هذا التصنيف تستمد من مصفوفة Matrix تدلنا على أى المصطلحات التى تحدث فى أى الوثائق (مصفوفة المصطلح/ الوثيقة)، ومن الواضح أن العملية العكسية، يمكن أيضاً أن تتم على أساس استخدام هذه البيانات فيمكن أن تتكون مجموعات الوثائق أو أقسام الوثائق على أساس المصطلحات التى تحتويها، وقد أظهر لنا سالتون بعض أنواع هذه الأقسام كما يلى: (Salton & Mc Gill,M., 1983)

(أ) المضلع The clique



وفى هذا المضلع فإن جميع المواد من أ ــ هــ ترتبط مع بعضها بقوة :

(ب) النجمة The star



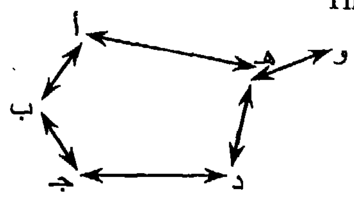
وفى هذه الحالة فإن مجموعة أ س هـ ك تعرف بحقيقة أن س، هـ، ك، ص كلها ذات ارتباط وثيق بالمادة أ بطريقة أو بأخرى.

(ج) الخيط The String

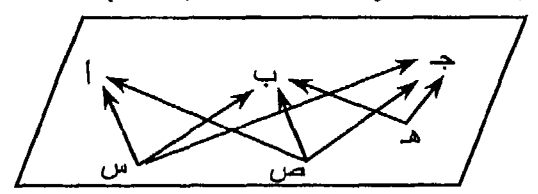
أ <-> ب <-> جـ <-> د <-> هـ

وفى هذه الحالة فإن ب ترتبط بشدة بـ أ، كما أن ج ترتبط بشدة بـ ب وهكذا حتى نصل إلى هـ وهى التى لا ترتبط بشدة إلا بمادة واحدة وهى د.

(د) النوع المختلط The clump



وهذه يمكن أن تتشكل على أساس أى واحدة من المعايير المتعددة، وبصفة عامة يرتبط كل عضو منه بالأعضاء الأخرى بالجماعة عند قيمة معينة أعلى من عتبة محددة، وهناك علاقات عديدة يمكن أن يتشكل من بينها الاستشهاد المتزامن (Swall, 1973...cited by Lancaster, 1991, p. 234).



المواد أ، ب، ج يمكن اعتبارها مكونة لقسم أو جماعة نظراً لأنه يتم الاستشهاد بها بواسطة كل من س، ص، أما من ناحية التزاوج الببليوجرافى Bibliographic Coupling جاء الاستشهاد المتزامن Co -citation يمكن أن يحدث بأطوال مختلفة، وفى الشكل السابق فإن المواد أ، ب، ج مرتبطة ببعضها ارتباطاً ضعيفاً نظراً لأن هناك مادتين فقط تستشهد بهما سوياً وكلما زادت المواد التى تستشهد بهما تزامنياً،كان الارتباط أقوى، ويلاحظ هنا أن الأقسام Classes التى تتشكل على أساس ارتباطات الاستشهادات لها بعض المزايا على الأقسام التى تتشكل من

خلال التكشيف الموضوعى التقليدى، ذلك لأنها مستقلة عن اللغة وعن المصطلحات المتغيرة، فقد يتغير اسم مرض معين أكثر من مرة فى زمن معين، ولكن ليس هناك ما يعوق القيام بالبحث عن هذا المرض فى كشاف الاستشهادات خاصة إذا كانت ورقة البحث التى قامت بتحديد هذا المرض معروفة للباحث، ويتم الاستشهاد بها بكثرة، أما مبدأ التزاوج الببليوجرافى فيمكن استخدامه فى ربط أوراق البحوث بلغات مختلفة تماماً وعلى سبيل المثال يمكن التعرف على أوراق بحوث منشورة باللغة الروسية أو الصينية والتى يوجد بها تزاوج قوى بورقة البحث الإنجليزية، وبالمثل يمكن أن يكون هناك قسم أو جماعة عن أوراق البحوث ذات الاستشهادات المتزامنة Co - cited وهذه يمكن أن تشمل مواد بلغات مختلفة، والشيء المهم بالطبع هو أن الأقسام التى تتكون بالاستشهادات المتزامنة تتغير مع الوقت، حيث ستوجد علاقات داخلية جديدة بناء على نتائج البحوث التى يعدها باحثون آخرون (Lancaster, F., 1991, p. 235) .

الفصل الخامس

تقييم الكشافات والتكشيف

مقدمة

تبدأ الدراسة باستعراض العوامل التى تؤثر على نوعية التكشيف أو التى تحول بين الكشاف وجودته، ثم تتناول تقييم قاعدة المعلومات حيث يرتبط التكشيف والاحتياجات المعلوماتية بكفاءة قاعدة المعلومات. وتتناول الدراسة أيضاً تجربة الباحثين وايت وجريفت فى قياس نوعية التكشيف، وتأثر نوعية التكشيف وجودته بدرجة الانتظام فى التكشيف. وتنتهى الدراسة بعرض لمشكلات التكشيف خصوصاً تلك المتصلة بالمجالات المعرفية غير اللفظية.

أولاً : بعض العوامل التى قد تحول بين الكشاف والجودة :

تعتبر البحوث المتصلة بهذه العوامل فى الإنتاج الفكرى قليلة جداً، وستحاول هذه الدراسة استعراض بعض هذه العوامل كما جاءت فى التسعينيات، ومن الملاحظ أن معظم البحوث التى تهدف إلى التعرف على مدى جودة الكشاف والتكشيف تتوجه للمستفيدين للتعرف على مدى رضائهم عن الكشاف فى استرجاع المعلومات التى يطلبونها. ولكن هناك بعض الجوانب التى تحتاج إلى التقييم ولكنها لا تتطلب رأى المستفيدين مباشرة وذلك مثل الأخطاء التى قد تحدث عند اختيار الواصفات الموضوعية وغيرها من الأخطاء فضلاً عن الإهمال العام فى بناء الكشاف وعرضه على المستفيد .. ولا ننسى ونحن بصدد نوعية وجودة الكشاف المستخدمين أنفسهم الذين قد لا يعرفون كيف يستخدمون الكشافات والإفادة منها أى أن المستفيدين فى حاجة إلى تعلم كيفية الاستخدام لتحقيق نجاح أكبر فى الاسترجاع ... وفيما يلى بعض هذه الجوانب التى تقلل من جودة ونوعية الكشاف (Cleveland, D. B, 1990, 152).

أ – أخطاء اختيار الواصفات الدالة على الموضوعات:

هناك الأخطاء التى تأتى من الاختيار غير السليم للواصفات، وقد يختار المكشف المهمل بعض الكلمات التى تغير المعنى الأصلى .. ومهمة التحرير هى التقليل على قدر الإمكان من هذه الأخطاء .. ومع ذلك فيجب الاعتراف بأن اختيار

مصطلحات الكشاف الصحيحة لتمثيل المعلومات يعتبر أمراً عسيراً سواء فى التكشيف اليدوى أو الآلى، وهناك أخطاء الحذف أى حذف بعض المصطلحات الأساسية فى التكشيف وتصحيح ذلك ربما يستدعى من المحرر إعادة تكشيف الوثيقة من جديد. وقد تكون السياسة الموضوعة للتكشيف هى السبب فى الحذف السالف الذكر، فقد تنص القواعد التى يعمل على أساسها المكشف بضرورة إلغاء الخلفية التاريخية مما يسبب الشكوى من المستفيدين أحياناً ... وهذا يتطلب مراجعة السياسة التحريرية العامة.

ب – البحث عن الرؤوس العامة:

تقدم الكشافات الهجائية عادة مشكلات وصعوبات فى عملية البحث العام generic searching .. وعندما يتاح التخصيص الشديد Specificity فإن الباحث عليه أن يبحث فى خيط طويل من المداخل ..

ج – المصطلحات:

تقييم المصطلحات أساسى فى بناء المكنز، وبناء المكنز أساسى فى النشاط التكشيفى ونوعيته وجودته، وإذا ما رفض كل من المؤلفين والمستفيدين مصطلحات المكنز فإن ذلك سيضع المكشف فى مأذق .. وعلى كل حال فالكشاف الجيد هو الذى سينتج للمستفيدين المصطلحات التى يتوقعون وجودها.

د – الإرشاد الداخلى :

أى وجود ملاحظات وإحالات لإرشاد المستفيد وتوجيهه للبحث فى المكان الصحيح أو اختيار البدائل، وهناك من المحررين من يفضل وضع تعليمات مطبوعة عن كيفية استخدام الكشاف، على أن تكون هذه التعليمات واضحة ومحددة وتغطى النقاط الرئيسية فى البحث.

هـ – الدقة فى الإحالات:

إذا كان المستفيدون يعتمدون على هذه الإحالات للحصول على المعلومات التى يحتاجونها، فإن طبيعة ودقة وعدد هذه الإحالات سيعكس نوعية الكشاف.

و – الهجاء والترقيم :

الهجاء غير الصحيح يلقى على الكشاف وجودته تعتيماً وبالتالى سيقلل من ثقة المستفيد بالكشاف. ويحاول كل من المكشفين والمحررين تجنب أخطاء الهجاء، كما أن الترقيم (أي استعمال النقط والفواصل .. إلخ) قد يمثل شكلاً آخر من الأخطاء .. وأفضل علاج هنا هو بيان وتعريف المكشف بالقواعد التى تستخدم فى الترقيم (استخدام الشرطة، الفصلة....)، كما أن الترقيم يعتبر مشكلة رئيسية فى عملية الترتيب الهجائى خصوصاً مع وجود أعداد ورموز وعلامات ترقيم.

ز – التكاليف:

التكاليف عامل لا يمكن تجنبه فى إنتاج الكشافات الجيدة؛ وتشمل التكاليف هنا رواتب المكشفين المهنيين وغيرهم من المساعدين، وهناك تكاليف إعداد المكنز واستمرارية مراجعته ووضع قائمة استناد واستمرار مراجعتها حتى لا تتقادم..

ح – المعايير :

تعتبر معايير التكشيف والاستخلاص أساسية، ذلك لأنها تؤدى إلى انتظام العمل consistency فضلاً عن تكوين الخبرة والإفادة القصوى منها.

وقد نشرت بالولايات المتحدة معايير التكشيف تحت اسم Basic Criteria for indexes وآخر طبعة عام ١٩٨٤ بواسطة المعهد الأمريكى الوطنى للمعايير [ANSI] وقد جاء فى هذه النشرة أن الهدف منها ليس تعليم الممارسات التكشيفية الأساسية، ولكنها تخدم كمرشد للنتائج التي يراد تحقيقها. هذا وقد قام معهد المعايرة الأمريكى بنشر أكثر من ثلاثين نشرة معايير تناولت جوانب عديدة فى المكتبات والمعلومات ذات أهمية وفائدة للمكشفين والمستخلصين ومنها:

- مبادئ أساسية للكشافات.
- كتابة وبناء المستخلصات.
- إرشادات لبناء المكانز.
- تصحيح المسودات.
- المراجع البيبليوجرافية.

ط – السمعة:

تعتبر الطريقة التقليدية التى يستخدمها أمين المراجع لتقييم الكشافات ذات فائدة كذلك، لأنها تدور حول سلسلة من الأسئلة كما يلى :

- من هو الناشر (مستوى الناشر وسمعته العالية يأتى معها كشافات جيدة).
- ما هو نطاق الكشاف؟ أى هل يغطى الكشاف ما ينبغى تغطيته فى الموضوع؟
- هل هذا الكشاف يكرر الكشافات الأخرى المتوفرة؟
- هل يغطى هذا الكشاف فجوات فى الكشافات الأخرى؟
- هل العمق والشمول فى التكشيف كافيا للمستفيد.
- هل يصدر الكشاف فى وقت سريع وجارٍ.
- هل يستحق التكاليف التى تنفق عليه.

وأخيراً فينبغى أن نتقبل الحقيقة القائلة بأنه ليس هناك كشاف مثالى نظراً لوجود متغيرات عديدة أهمها تدخل الإنسان فى اختيار المصطلحات التى قد يكون فيها تحيز .. كما أن الكشاف الجيد هو الذى يسمح للمستفيد ــ بصفة عامة ــ أن يجد بسرعة وسهولة كل المعلومات المتصلة باحتياجاته المعلوماتية.

ثانياً : التعرف على التكشيف الجيد كما يراه لانكستر:

يذهب لانكستر (Lancaster, 1991, 77) إلى أن الحكم على نوعية التكشيف يعتمد إلى حد كبير على نتيجة الخبرة فى تشغيل نظام الاسترجاع وتقييمه، ذلك لأننا لا نستطيع أن نحكم على مجموعة من المصطلحات الكشفية التى تم تعيينها لوثيقة معينة ــ بأنها صحيحة أو غير صحيحة فى شكل تجريدى مطلق .. وإن كان من الممكن أن يقوم رئيس المكشفين بتحديد بعض الأخطاء فى هذا التمثيل للوثيقة قبل إضافتها لقاعدة البيانات .. أى أنه يقوم ببعض التحكم النوعى لعملية التكشيف .. والأخطاء التى سيجدها رئيس المكشفين يمكن أن تتخذ الأشكال التالية:

١- السياسة التى اتبعها المكشف خصوصاً تلك المتعلقة بشمولية التكشيف (exhaustivity).

٢- فشل المكشف فى استخدام عناصر المصطلحات بالطريقة التى يجب أن تستخدم بها (رأس موضوع رئيسى غير صحيح مثلاً).

٣- فشل المكشف فى استخدام مصطلح فى المستوى الصحيح للتخصيص..

٤- استخدام المكشف لمصطلح غير صحيح نظراً لعدم معرفته بالموضوع العلمى.

٥- حذف أو إهمال المكشف لبعض المصطلحات الهامة.

لقد اعتمد لانكستر فى تقييمه على الفطنة والتحليل الذاتى، وليس على الأدلة المحققة والتجارب، ونتيجة اجتهاده، فقد قسم العوامل التى تؤثر على نوعية التكشيف إلى خمسة عوامل، بعضها يعود للمكشف والآخر للوثيقة والثالث للمصطلحات والرابع لعملية التكشيف نفسها وأخيراً العوامل البيئية وذلك كما يلى:

عوامل تعود للمكشف	عوامل تعود للوثيقة
- المعرفة الموضوعية - معرفة احتياجات المستفيد - الخبرة - التركيز - القدرة القرائية والفهم	- المادة الموضوعية - التعقد الموضوعى - اللغة - طول الوثيقة - طريقة تقديمها وتلخيصها
عوامل تعود للمصطلحات	**عوامل تعود لعملية التكشيف**
- التخصيص - الغموض أو عدم الدقة - نوعية المصطلحات كمداخل - نوعية تركيب تلك المصطلحات - توفر الأدوات المساعدة	- نوع التكشيف - القواعد والتعليمات - الإنتاجية المطلوبة - شمولية التكشيف
	العوامل البيئية
	- الحرارة والبرودة - الإضاءة - الضجة

أ - العوامل التى تتصل بالمكشف:

يجب أن يكون لدى المكشفين دراية بالمادة الموضوعية التى يتعاملون معها مع ضرورة فهم مصطلحات تلك المادة الموضوعية، على الرغم من أنه ليس من

الضرورة أن يكون هؤلاء المكشفين خبراء فى الموضوع .. أى أن الدراسة هنا ضرورية للتمييز بين مصطلحين فى الموضوع ذلك لأن المكشف مثلاً قد يعين مصطلحين لتمثيل الوثيقة بينما الصحيح الاكتفاء بمصطلح واحد فقط وقد أطلق الباحث لوكوبولس (Loukopulos, 1966) على ذلك عدم قدرة المكشف على اتخاذ القرار.

هذا وقام دياداتوا (Diodato, 1981) بدراسة مدى الانتظام Consistency فى اختيار المصطلحات بالنسبة لثلاثة مجموعات من المكشفين وهم المؤلفون والمكشفون المهنيون وقراء أوراق بحوث الرياضيات .. وقد أثبتت هذه الدراسات أن المؤلفين يقومون بتكشيف ذى دقة أفضل better Precision ولكنه منخفض الاستدعاء low recall ولكن دراسة رشيد (Rasheed, 1989) التى قارن فيها بين المصطلحات التى قام بتعيينها مؤلفو المقالات الطبية بالمصطلحات التى قام بتعيينها مكشفو نظام الميدلرز Medlars ، وتبين له أن المكشفين قاموا بتعيين مصطلحات أكثر وأن المصطلحات التى قاموا باستخدامها كانت أكثر تخصيصاً من تلك التى استخدمها المؤلفون.

إن معرفة اهتمامات المستفيدين من قاعدة المعلومات يعتبر ذا أهمية خاصة نظراً لأن التكشيف "الجيد" يجب أن يفصّل طبقاً لاحتياجات مجتمع معين كلما أمكن ذلك ويدخل فى ذلك أيضاً خبرة المكشف ومقدرته على التركيز والقراءة والفهم السريع لمضمون الوثيقة فضلاً عن استمتاعه بعمله.

ب- العوامل التى تتصل بالوثيقة :

هناك بعض الموضوعات عسيرة الفهم أكثر من غيرها، فالنظرية تكون عادة أكثر صعوبة من الممارسات، ويتصل بهذا الجانب كذلك درجة مضاهاة المادة الموضوعية للوثيقة ومعرفة اهتمامات المكشف. وتأتى لغة الوثيقة من بين العوامل التى تؤثر على نوعية التكشيف، فبعض المؤلفين يقومون بالتعبير عن أفكارهم أو نتائج بحوثهم بطريقة أكثر وضوحاً من سواهم وبالتالى يجعلون مهمة المكشف أكثر يسراً..

ج – العوامل المتصلة بالمصطلحات :

هذه العوامل يمكن أن تؤثر على نوعية التكشيف، فكلما كانت المصطلحات أكثر تخصيصاً كلما كان التعبير عن المعنى أكثر وضوحاً .. كما أن إضافة بعض العناصر الأخرى كرؤوس الموضوعات الفرعية أو مشيرات الدور ستزيد من تخصيص المصطلح ولعل ذلك قد يعقد من مهمة المكشف أحياناً، فالقول السائد هنا هو أن يتبع المكشف قاعدة التخصيص، ولكن ألا يكون مسرفاً فى عمق التخصيص.

د – عوامل تعود لعملية التكشيف:

يتوقع من المكشفين أن يقوموا بعملهم بدرجة أكثر فاعلية عندما تكون تحت أيديهم قواعد وتعليمات دقيقة، كما أن الإنتاجية التى قد تطلبها بعض الهيئات من المكشفين قد تفرض عليهم بعض الضغط مما يؤدى إلى أخطاء .. كما أن التكشيف الشامل exhaustive قد يتطلب وقتا أطول من التكشيف الانتقائى.

هـ- وأخيراً فالتكشيف يحتاج إلى تركيز فى التفكير وظروف بيئية مناسبة.

ثالثاً : تقييم قاعدة البيانات :

يمكن الحكم على جودة التكشيف إذا أتاح للباحثين تحديد المواد التى يحتاجونها دون ضرورة النظر إلى مواد أخرى لا يحتاجونها. كما أن قاعدة البيانات الببليوجرافية لا يمكن تقييمها فى عزلة عن العوامل الأخرى ولكن يتم الحكم عليها بالنسبة لقيمتها فى الاستجابة لمختلف الاحتياجات المعلوماتية.

هذا ومن العسير التعرف على نوعية التكشيف كموضوع بذاته، ولكننا نتعرف عليه عند تقييم نظام استرجاع المعلومات، والذى يعتبر التكشيف واحداً فقط من العوامل المؤثرة العديدة على الأداء الكلى وفى الواقع فهناك إنتاج فكرى كثير على انتظام Consistency عملية التكشيف وليس على نوعيته.

هذا وقد قام الباحث لانكستر (Lancaster, F., 1991, 116) بربط التكشيف والاحتياجات المعلوماتية بكفاءة قاعدة البيانات. وقد وضع المعايير الأربعة التالية لهذا التقييم:

أ – التغطية: أى ما مدى تغطية قاعدة البيانات للإنتاج الفكرى عن الموضوع والمنشور خلال فترة زمنية معينة

يعتبر تقييم قاعدة البيانات شبيه إلى حد كبير بتقييم اكتمال المقتنيات بالمكتبة بالنسبة لموضوع معين، وإحدى طرق تقييم مجموعات المكتبة فى موضوع معين هو العثور على ببليوجرافيات موثوق بها فى الموضوع ثم مضاهاتها بالمجموعات بالمكتبة .. وبالمثل يمكن تقييم الكشاف الطبى Index Medicus مثلاً عن طريق العثور على قائمة ببليوجرافية شاملة عن موضوع طبى معين لفترة محددة، حيث يتم مضاهاة مداخل الببليوجرافيا على كشاف المؤلفين فى الكشاف الطبى للتعرف على المواد الداخلة فى الكشاف والمواد غير الداخلة فيه ... ونتيجة لهذه المضاهاة يمكن القول مثلاً بأن الكشاف الطبى يحتوى على (٨٤%) من الإنتاج الفكرى فى هذا الموضوع .. مع الأخذ فى الاعتبار سياسة الكشاف الطبى بالنسبة لتغطية الدوريات دون الكتب.

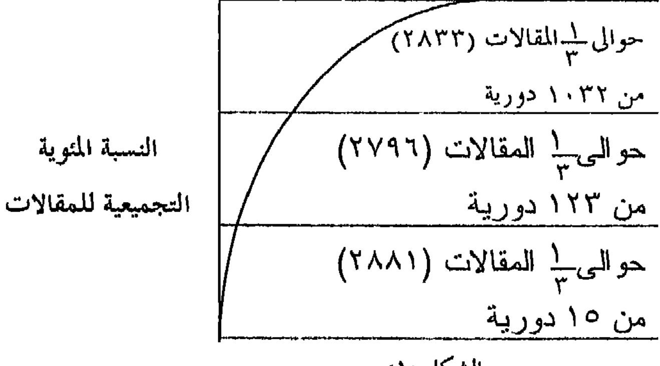
وهذا الأسلوب لا يخلو من صعوبات ومشكلات فالعثور على ببليوجرافيا شاملة موثوق بها أمر عسير، فضلاً عن ضرورة التعرف على طريقة تجميع تلك الببليوجرافيا، فقد تكون هى نفسها معتمدة على مخرجات الكشاف الطبى نفسه وبالتالى فالتقييم هنا لا طائل من ورائه.

وقد لا نحتاج إلى ببليوجرافيا شاملة، فالعينة الممثلة لموضوع معين قد تكون كافية. وإحدى الطرق للحصول على تلك العينة هو استخدام وإحدى قواعد البيانات كمصدر للمواد التى يمكن بواسطتها تقييم تغطية قاعدة بيانات أخرى .. وعلى سبيل المثال فلمعرفة درجة اكتمال تغطية الكشاف الهندسى فى موضوع شبه الموصلات Semiconductors يمكن اللجوء إلى المستخلصات الفيزيائية للتعرف على حوالى (٢٠٠) مادة قامت بتكشيفها تلك الخدمة، ثم استخدام هذه المجموعة لتقدير مدى تغطية الكشاف الهندسى لهذا الموضوع .. وواضح أن (٢٠٠) مادة ليست هى كل المنشور عن الموضوع ولكن يمكن اعتبارها عينة ممثلة ثم معرفة النسبة المئوية للتغطية إذا كانت هناك مثلاً (١٤٠) مادة فى الكشاف الهندسى فى هذه المجموعة فتصبح النسبة (٧١%) ١٤٢÷ ٢٠٠ والعكس صحيح أى تقييم المستخلصات الفيزيائية بمضاهاتها بعينه من الكشاف الهندسى .. وفى هذه الحالة

يمكن التعرف على التداخل Overlap بين القاعدتين .. كما أن هذا المقياس يصلح فى القواعد المطبوعة والمحسبة آخذين فى الاعتبار تواريخ المطبوعات.

ويهمنا فى التقييم معرفة التغطية لأن بعض مراكز المعلومات قد تحتاج إلى معرفة مدى تغطية قاعدة بيانات معينة (مثلاً المستخلصات الكيميائية) لمجال تخصصها. أم أنها فى حاجة إلى الاعتماد على عدة بيانات أخرى للتغطية الكاملة .. كما أن منتج قاعدة البيانات فى حاجة كذلك لمعرفة مدى تغطيته لموضوع معين وأشكال الوثائق التى يتعامل منها.

وهناك موضوع هام بالنسبة لتقييم التغطية وهو موضوع التشتت Scatter فيذكر لانكستر (Lancaster, F. 1990, 121) أن كل الإنتاج الفكرى عن مرض الإيدز عام ١٩٨٢ طبقاً لبحث قاعدة بيانات ميدلاين كان بثلاث لغات وفى ١٤ دورية فقط ولكنه عام ١٩٨٧ كان بخمس وعشرين لغة واسهمت أربع وخمسين دولة فى الإنتاج الفكرى وظهر ذلك فى (١٠٢٠٠) دورية علمة ومعنى ذلك أنه كلما نما الإنتاج الفكرى كلما زاد انتشاره (بلاد أكثر، لغات أكثر، دوريات أكثر، أشكال وثائق أكثر..) وبالتالى يصبح من العسير تحديد وتجميع وتنظيم هذا الإنتاج المتخصص .. وقد أوضح برادفورد ظاهرة الانتشار هذه بالنسبة للإنتاج الفكرى لمرض الإيدز والذى يظهر فى المنحنى التالى (الشكل١) والذى يمكن أن تفيد منه مراكز المعلومات فى اقتناء الدوريات الأكثر إنتاجاً وهى (١٥) دورية فقط.



الشكل (١)

برادفورد وظاهرة التشتت

النسبة المئوية التجميعية للدوريات

ب – الاسترجاعية Retrievability

أى ما هو مقدار ما يمكن استرجاعه من المواد فى قاعدة البيانات باستخدام استراتيجية بحث مناسبة. إذا كانت قاعدة البيانات تحتوى على عدد (N) من مواد موضوع معين (وهذه يمكن معرفتها عن طريق دراسة التغطية)، فكم عدد المواد التى يمكن استرجاعها عند بحث هذه القاعدة من بين المواد الموجودة فيها؟

والإجابة على هذا السؤال يمكن معرفتها عن طريق دراسة التغطية .. لنفترض أننا نريد دراسة التغطية والاسترجاعية لعدد من الموضوعات التى تحتويها قاعدة البيانات الزراعية أجريكولا Agricola .. فإذا كنا نعرف مقدماً أن هناك (٨٠) مادة عن الحشرات فى قاعدة بيانات أجريكولا .. وأن الباحث استرجع من هذه عدد (٦٠) مادة فقط، فإن الاستدعاء Recall يصبح ٧٥% فقط .

وواضح أن هذا الاختيار، لا يتضمن اختبار قاعدة البيانات وتكشيفها فحسب ولكنه يتضمن قدرة الباحث وخبرته كذلك والتى قد تختلف من باحث إلى آخر وبالتالى ستختلف النتائج، كما يلاحظ كذلك أن الاسترجاع هنا (الاستدعاء) يتم الحكم عليه من أساس المواد الصالحة والمعروفة سلفا والموجودة فى قاعدة البيانات. وهناك طريقة بديلة لدراسة استرجاعية المواد من قاعدة البيانات وهذه تتضمن نوعاً من المحاكاة Simulation.

وقد قام أولبريت (Albright, 1979) بدراسة تفصيلية فى هذا الاتجاه باستخدام الكشاف الطبى Index Medicus. وقد تبين له بناء على بحوث المحاكاة التى قام بها بالنسبة لعشرة موضوعات مختلفة أنه فى المتوسط، يجب استشارة حوالى (٤٤) مصطلحا مختلفاً لاسترجاع جميع المواد المعروفة بأنها صالحة relevant لموضوع معين. ويلاحظ هنا أن بعض هذه المصطلحات كانت مترابطة أى موصولة بنظام المصطلحات (عن طريق العلاقات الهرمية hierachical أو الإحالات Cross-Reference)، إلا أن العديد من هذه المصطلحات لم يكن له هذا الترابط.

ويلاحظ أنه كما تنتشر مقالات الدوريات على عناوين الدوريات، فإن المواد الخاصة بموضوع معين موجود فى قاعدة البيانات يمكن أن تنتشر أيضاً بالنسبة للمصطلحات المختلفة، ويوضح ذلك الشكل التالى:

النسبة المئوية التجميعية للمواد المسترجعة	
	مواد من المحتمل ألا يتم العثور عليها
	مواد يمكن العثور عليها من خلال إضافة مصطلحات قريبة
	مواد تم العثور عليها تحت عدد صغير من المصطلحات "الواضحة"
	النسبة المئوية التجميعية للمصطلحات

الشكل (٢) تشتت المواد تحت مصطلحات التكشيف

ويظهر لنا من الشكل (٢) أنه بالنسبة لأى موضوع محدد، فهناك نسبة مئوية عالية نسبياً للمواد الصالحة relevant يمكن أن توجد تحت عدد صغير من المصطلحات الواضحة .. ويمكن عند إضافة عدد من المصطلحات المرتبطة، فإن الاسترجاع سيرتفع مثلاً إلى (٨٠-٩٠%) .. ويبقى فى هذه الحالة الافتراضية نسبة مئوية (١٠-٢٠%) من المواد الموجودة والتى لن يستطيع الباحث العثور عليها.

ج - التنبؤ Predictability :

أى كيفية قيام المستفيد بالحكم على المواد المفيدة فى قاعدة المعلومات، وقد أوضحت لنا المناقشات عن تقييم الاسترجاعية أن هناك افتراضاً أساسياً يتمثل فى إمكانية التعرف على المادة الصالحة relevant من المعلومات المتعلقة بهذه المادة فى قاعدة البيانات .. وهذه المعلومات يمكن أن تشمل :

- عنوان المادة.
- العنوان بالإضافة إلى قائمة مصطلحات الكشاف.
- العنوان بالإضافة إلى المستخلص.
- العنوان بالإضافة إلى مصطلحات من المستخلص.

ويذهب لانكستر (Lancaster, 1991, p. 131) إلى أنه كلما كان التمثيل للوثيقة أطول، كان هناك مفاتيح أكثر عن مدى أهمية المادة للمستفيد - فالعنوان وحده قد

يكون مفيداً ولكن ضم العنوان إلى المصطلحات الكشفية يمكن أن يكون أداة قوية لبيان ما تقوله الوثيقة .. وتعتبر المستخلصات بالطبع أفضل المشيرات إلى المحتوى .. وإن كان ذلك ليس حتمياً فى مختلف التجارب.

د – السرعة : Timeliness :

ويتمثل هذا العامل فى مدى سرعة استرجاع المواد المنشورة حديثاً، ومدى ما قد يؤدى إليه التكشيف والاستخلاص إلى موقف تكون فيه المواد المسترجعة متأخرة قديمة. فالحداثة هى قياس السرعة التى يتم بها إدخال المطبوعات الجديدة فى خدمات التكشيف والاستخلاص ولكن تأخر وصول المطبوعات لخدمة التكشيف والاستخلاص قد يكون لأسباب عديدة منها طول الفترة بين صدور الدورية واستلامها فى خدمة التكشيف، وقد يكون التأخير لأسباب جغرافية أو اقتصادية، كما قد يكون من العسير تحديد مكان أوراق المؤتمرات وقد يكون الأمر أكثر عسراً للحصول عليها .. والوثائق التى تصدر بغير اللغة الإنجليزية قد تحتاج إلى فترة لترجمتها وقد يكون سبب التأخير لعدم كفاءة النظام ذاته. ومن وجهة نظر المقيم فإن السرعة تمثل أحد المعايير الهامة لفاعلية الخدمة.

وإذا كانت التغطية والاسترجاعية والتنبؤ والسرعة هى المعايير التى وضعها لانكستر لتقييم قواعد البيانات والكشافات، فإن هناك معايير أخرى عديدة كتلك التى تتمثل فى **الكفاية** Adequacy (وتغطى هذه الصفة جوانب عديدة بما فيها التغطية والمصطلحات المستخدمة فى التكشيف والشمول والانتظام) وكذلك العمومية (وهذه تتصل أساساً بتنوع البحوث التى يمكن القيام بها)، **سهولة الاستخدام** والحالية Currency (أى إلى أى مدى تستحدث الأدوات) ثم بعد ذلك التكاليف.

رابعاً: تجربة وايت وجريفيت فى قياس نوعية التكشيف وتعليق الباحثين عليها:

قام كل وايت وجريفيت (White, H,1987) باختبارات عملية للحكم على جودة ونوعية التكشيف الموضوعى فى عدة قواعد معلومات ببليوجرافية على الخط المباشر والمقارنة بينها .. وتتمثل هذه الاختبارات فى عدد (١٨) مجموعة من الوثائق، وتم اختيارها من الإنتاج الفكرى الطبى السلوكى، مع بعض المصطلحات المأخوذة من ميدلاين Medline ، سايكوانفو Psych-Info بيوزيس Biosis والمستخلصات الطبية Excerpta Medica .. وقد تضمن كل اختبار مجموعة

Cluster من حوالى خمس وثائق لها ارتباط فيما بينها بالنسبة للمادة الموضوعية (بناء على بعض الأسس التى وضعها الباحثان) ثم قام الباحثان باسترجاع واصفاتها من قاعدتين للمعلومات على الأقل ثم قاما بإعداد جداول بالرقم المتوسط للواصفات المستخدمة فى الوثائق، وعدد الواصفات المستخدمة فى جميع الوثائق فى المجموعة أو لمعظمها، وكذلك الندرة النسبية للواصفات المستخدمة، وقد تولدت إحصاءات مقارنة عن كيفية ربط كل قاعدة بيانات الوثائق القريبة من بعضها وكيفية التمييز العريض والدقيق بين الوثائق. وقد أدت هذه الاختبارات ــ كما يقول الباحثان ــ إلى اكتساب نظرات ثاقبة نوعية فى مدى التعبير expressiveness والصلة بالموضوع Pertinence بالنسبة للمصطلحات الكشفية المتوفرة.

ويلاحظ أن هذه الاختبارات لم تتم بالطرق السابقة المتمثلة فى اختبارات الاسترجاع التى تشمل أساساً طلبات المستفيدين ثم القيام بالحكم عليها من ناحية الصلاحية ..Relevance (Spank, J, 1981) وبالمقارنة فإن هذه التجارب قد قصد بها أن يتبين مديرو قواعد البيانات أو المستفيدون ماذا قام به المكشفون عند تطبيق المصطلحات المحكومة .. وبالتالى فإن هذه التجارب توسع من إدراكنا بالنسبة لكيفية استخدام المكشفين للغة عندما يحاولون التقاط المحتوى بمكنز معين وببعض القواعد الإرشادية .. وبالتالى فهذه الطريقة لا تؤدى إلى نقد نظم استرجاع الوثائق الكاملة ولكنها تؤدى إلى نقد مكونات التكشيف ذى المصطلحات المحكومة.

٤/١ أبعاد التكشيف وتطبيق المنهجية البحثية :

لقد أدى تطور الخدمات الببليوجرافية على الخط المباشر إلى إنشاء إمكانيات بحث جديدة من خلال استرتيجيات البحث (And/or/and not) .. وقد قام الباحثون بتحديد ثلاثة أبعاد للتكشيف وهى: (أ) وصل الوثائق المرتبطة ببعضها. (ب) التمييز العريض بين هذه المجموعات المرتبطة داخل الملف الكامل (ج) التمييز الدقيق بين الوثائق الفردية.

وكانت نقطة البداية لوضع المنهجية هى تحديد مجموعات الوثائق المعروفة ــ على أساس مسبق ــ متشابهة فى محتواها، وهذه الوثائق تسمى مجموعات الوثائق الحاسمة Critical document clusters نظرا لأنها تقدم لنا معياراً للقياسات التالية، وكل واحدة من الوثائق من مجموعة معينة يتم بحثها فى قاعدة البيانات

بواسطة المؤلف وغير ذلك من المصطلحات .. ثم يتم طباعة وجدولة واصفات هذه الوثيقة. ومن هذه الجداول يمكن حساب القياسات كما يلى:

- وصل الوثائق المرتبطة Linking related documents

هذا المقياس يدل على عدد الواصفات التى تربط جميع أو معظم الوثائق فى مجموعة واحدة والسؤال هنا هو: هل قام المكشف بتعيين نفس المصطلح إلى نصف (أو أكثر من النصف) للوثائق المرتبطة وبالتالى التقاط نقطة التشابه الموضوعى بينهم؟

- **التمييز العريض بين المجموعات الأصغر من الوثائق المرتبطة مع بعضها:**

وهذا المقياس يدل على المدى الذى تكون فيه الواصفات التى تربط المجموعة غير كثيرة نسبياً فى الاستخدام، أما الواصفات المستخدمة كثيراً جداً فلا تؤدى إلى التمييز الجيد، فالواصف المستخدم لجميع الوثائق فى قاعدة البيانات سوف لا يؤدى إلى عملية التمييز نهائياً وبالتالى فلا يعتبر ذا فائدة، وهناك أيضاً المصطلح الذى يستخدم نادراً جداً، وبالتالى فهو يقوم بالتمييز الجيد تماماً، وعندما يتم إضافته للمصطلحات الأخرى فإنه يؤدى إلى استرجاع وثائق قليلة أو لا يسترجع أى وثيقة.

هذا وعدد مرات تكرار الواصف يدلنا على مدى كثرة استخدامه فى قاعدة البيانات، والبديل لهذا الرقم يدلنا على الندرة النسبية لاستخدامه .. وقد قام الباحثان بتحويل الأرقام البديلة إلى كشاف يتراوح بين صفر إلى واحد، وكلما كان الكشاف أعلى كانت هناك ندرة فى استخدام المصطلح (وكما يذهب سبانك Spank فإن هذا المقياس يدلنا على التخصيص Specificity أو تخصيص التكشيف عن طريق حساب استخدام المصطلح عبر مجموعة الوثائق وهذا من الناحية الإحصائية وليس الدلالية).

- **التمييز الدقيق بين الوثائق الفردية:**

وهو قياس يدلنا على متوسط عدد الواصفات المعينة لكل وثيقة فى المجموعة Cluster [ويعبر عن هذا المفهوم بمصطلح الشمولية Exhaustisity] ومن الواضح أن ذلك يمكن أن ينتج عند الأخذ فى الاعتبار ندرة المصطلحات .. وقد قام الباحثان بالتعرف على قواعد البيانات التى لديها متوسطات عالية نسبياً، على

افتراض أن كل مصطلح إضافى يعبر عن محتوى الوثيقة بطريقة أكثر شمولاً وبالتالى فإنه يتيح الاسترجاع على أبعاد مفهومية أكثر.

ويلاحظ هنا خاصية هامة فى منهجية هذه الدراسة، وهى الحصول على مقاييس مقارنة على قواعد البيانات المختلفة، أى أن يقوم باحث بدراسة نفس وثائق المجموعة فى اثنين (أو أكثر) من قواعد البيانات، ثم يقوم بطباعة المخرجات ووضعها فى جداول وذلك بالنسبة للواصفات المعينة فى كل قاعدة، ثم يحسب نفس المقياس لكل قاعدة.

وهذا الاختبار المباشر للافتراض بأن الوثائق المجمعة متشابهة فى المحتوى ولكن الأكثر أهمية هو أن مقارنة المقاييس تقيّم مباشرة نوعية التكشيف فى قواعد البيانات المختلفة.

وفى التطبيق الدقيق فإن المسلك المقارن محدود بالطبع باستخدام هذه المنهجية لمجموعات نفس الوثائق المشتركة فى عدة بيانات، وكل واحدة منها يجب أن تغطى نفس الموضوع وتكشف نفس الوثائق.

وأخيراً فإن العرض الموازى للواصفات الخاصة بنفس الوثائق من قاعدتين للبيانات، يوضح لنا فحصاً نوعياً للتكشيف كتعبير لغوى، فهو يعطينا بؤرة جادة لإمكانيات الإضافة والتغيير فى المصطلحات داخل قواعد البيانات وكذلك للتعديلات التى تتم فى ممارسة التكشيف، حتى ولو لم تتم المقاييس الكمية.

وباختصار فتجربة وايت وجريفت تختبر المنهجية المقترحة أعلاه على (١٨) مجموعة وثائق فى العلوم الطبية السلوكية (MBS) .. ولكل مجموعة تتم مقارنة تكشيف الميدلاين MEDLINE مع قاعدة بيانات أخرى من القواعد التالية Psyc Info, Biosis Previews, and Excerpta. Medica.

٢/٤ ملخص النتائج الرئيسية:

١- كانت قاعدة بيانات ميدلاين MEDLINE أعلى قاعدة بالنسبة لمتوسط عدد المصطلحات المعينة وفى عدد المصطلحات الشاملة المميزة Spanning/ Disciminationg (S/D). والتكشيف المستخدم هنا يعكس انتظاماً فى

المستوى العالى سواء على المقياس الكمى أو بالنسبة لمدى تعبير المصطلح expressiveness.

٢- قامت قاعدة بيانات بيوزيز Biosis Previews بتعيين مصطلحات أقل قليلاً من الميدلاين، ولكنها زادت عن الميدلاين فى واحدة من أربعة مجموعات وضعت للمقارنة ويلاحظ أن كثيرا من مصطلحات بيوزيز شملت المجموعات المختلفة، ولكن جميعها قد استخدمت بكثرة فى قاعدة البيانات وبالتالى لم تعد صالحة أن تكون مميزة بدرجة عالية أى أنه لم توجد مصطلحات شاملة مميزة S/D . والكثير من مصطلحات بيوزيز كانت عامة ومجردة وينقصها التعبير expressiveness .

٣- قامت المستخلصات الطبية Excerpta Medica بتعيين مصطلحات أقل قليلاً من المستخدمة فى الميدلاين، ومع ذلك فإن المستخلصات الطبية كانت تحتوى على مصطلحات شاملة/ مميزة (S/D) أقل .. ويبدو ذلك غريباً نظراً لأن التكشيف فى المستخلصات الطبية معبر بوضوح .. كما أظهرت المستخلصات الطبية تغطية قوية لأسماء الأدوية.

٤- استخدمت قاعدة بيانات علم النفس Psyc Info مصطلحات لكل وثيقة أقل كثيراً من تلك المستخدمة فى الميدلاين .. ومع ذلك فقد كانت هذه القاعدة هى التالية مباشرة لقاعدة الميدلاين بالنسبة لاستخدام المصطلحات الشاملة/ المميزة (S/D)، كما ولدت هذه القاعدة عدداً لا بأس به من المصطلحات المرشحة لإدخالها ضمن القوائم المعيارية.

٣/٤ تعليق لانكستر :

ويعلق لانكستر على نفس هذه التجربة (Lancaster, 1991, 84) حيث يطلق عليها تقييم التكشيف خارج سياق نظام الاسترجاع الذى تحدث فيه .. أى أن دراسة وايت وجريفث (White, H, 1987) هى دراسة لنظام التكشيف من الخارج وذلك عن طريق تكوين مجموعة من الوثائق "المتشابهة فى المحتوى" ثم يتم استخدام مجموعات من هذا النوع كأساس للتقييم .. والباحثان ينظران إلى ثلاث صفات

مميزة لمصطلحات الكشاف التى تعين للمواد فى المجموعة داخل قاعدة معلومات معينة حيث يعرضهم لانكستر كما يلى مع التعليق:

أ - المدى الذى تربط فيه المصطلحات، المواد المتعلقة .. والمقياس الواضح لذلك هو عدد المصطلحات التى تم استخدامها لجميع أو لمعظم المواد فى المجموعة، ويمكن اعتبار المواد مرتبطة ببعضها بشدة إذا كان هناك بعض المصطلحات الموضوعية التى استخدمت لجميع المواد.

ب - المدى الذى يمكن للمصطلحات أن تميز بين هذه المجموعات داخل قاعدة البيانات والمقياس الأكثر وضوحاً هو عدد مرات تكرار المصطلحات التى تستخدم بالنسبة لمعظم الوثائق فى المجموعة داخل قاعدة البيانات ككل .. ويلاحظ هنا أن المصطلحات الشائعة جداَ ليست مصطلحات جيدة للتمييز بين الوثائق .. فعلى سبيل المثال ففى الميدلاين Medline يمكن للمصطلح إنسانى human أن ينطبق على كل مادة فى المجموعة ولكنه ذا قيمة ضعيفة فى تمييز هذه المجموعة عن الآخرين، نظراً لأنه نادراً جداً فى قاعدة البيانات ككل، وستكون مفيدة فى البحوث المتخصصة جداً، ولكنها ذات فائدة قليلة فى التعرف على المجموعات الأكبر نسبياً.

ج - المدى الذى يمكن فيه للمصطلحات أن تميز بين وثيقة وأخرى: الندرة تعتبر مقياساً يمكن تطبيقه هنا كذلك، فضلاً عن شمولية التكشيف فهناك مصطلح يمكن تطبيقه على جميع المواد فى المجموعة ولكنه لا يستطيع أن يميز بين أعضائه، وكلما كانت هناك مصطلحات إضافية أكثر تعيّن لكل عضو، أمكن تحديد الفروق الفردية بطريقة أفضل.

وللنظر إلى النوعية بهذه الطريقة فيجب أن ينشأ أولاً مجموعات الاختبار، ثم استرجاع تسجيلات لأعضاء كل مجموعة من قاعدة بيانات ثم دراسة صفات المصطلحات التى يتم تعيينها. وقد استخدم وايت وجريف هذا الأسلوب لمقارنة تكشيف مجموعات الاختبار فى قواعد بيانات مختلفة، فقد تمت مقارنة الميدلاين Medline مع سيكوانفو Psyc Info، مراجعات بيوزيز Biosis Preview والمستخلصات الطبية Excerpta Medica. ومقارنة قواعد البيانات بهذه الطريقة يعتبر مراجعة على الافتراض القائل بأن مواد مجموعة الاختبار متشابهة فى

المحتوى، وقد استخدم وايت وجريفت، المصاحبة الاستشهادية Cocitation كأساس لإنشاء مجموعات الاختبار، على الرغم من أن الطرق الأخرى، بما فى ذلك المزاوجة الوراقية يمكن استخدامها كذلك.

ويعلق الباحث لانكستر (Lancaster, 1991, 85) على هذه التجارب بقوله إن قيمة هذا العمل محدودة نظراً لأن التجمعات Clusters المستخدمة صغيرة جداً (تتراوح بين ثلاثة إلى ثمانية مواد)، كما أن صحة validity هذه الطريقة كاختبار للتكشيف الإنسانى تعتمد كلية على قبول الباحث لتجمع المصاحبة الاستشهادية كمعيار مناسب.. كما يرى لانكستر أن مراجعات "النوعية" يمكن القيام بها بطريقة أكثر يسراً: فمجموعات المواد التى يتم تحديدها بواسطة مصطلح أو مصطلحات معينة يمكن استرجاعها من عدة قواعد بيانات، ثم يتم مقارنة تكشيفها دون استخدام المصاحبة الاستشهادية كمعيار.

٤/٤ بعض النتائج العامة

تشير نتائج العديد من البحوث والاختبارات لقواعد المعلومات الإلكترونية إلى ما يلى:

أ – إدخال اللغة الطبيعية فى استراتيجيات البحث يؤدى إلى تحسينات ملموسة فى الاستدعاء Recall بالمقارنة باستخدام المصطلحات المحكومة وحدها.

ب– يجب اعتبار واسترجاع الاستشهادات Citation retrieval كملازم هام للاسترجاع بواسطة المصطلحات ذلك لأن إضافة مصطلحات متعلقة يمكن أن تكون مفيدة باستخدام مدخل الاستشهادات.

ج – ليس هناك قاعدة معلومات يمكن أن تزودنا بمفردها بتغطية شاملة كاملة لإنتاج فكرى معقد ومتعدد الارتباطات.

هذا وهناك بعض الملاحظات الإضافية فى هذا الصدد، من بينها أن أحد العوامل التى لها تأثير على نوعية التكشيف هو توفير الأدوات المرجعية المناسبة للمكشف وقد اقترح بيكول (Bakewell, 1987) حوالى مائتى كتاب مرجعى ذوى قيمة وأهمية للمكشفين.

خامساً: نوعية التكشيف وخاصية الانتظام: Quality and consistency

نوعية التكشيف ليست هى درجة انتظامه على الرغم من ترابطهما، وقد قام ليونارد (Leonard, 1975) بالتعرف على علاقة النوعية بالانتظام فى التكشيف وقد اعتبر ليونارد أن "النوعية" "فى إطار فاعلية الاسترجاع تعنى القدرة على استرجاع المواد المطلوبة وتجنب المواد غير المطلوبة، وقد تبين للباحث أن هناك علاقة "متوسطة إلى قوية" بين الانتظام وفاعلية الاسترجاع، مع علاقة إيجابية واضحة بين الانتظام ونسبة الاستدعاء. Recall ratio .. أى هناك علاقة إيجابية بين الانتظام ونوعية التكشيف على اعتبار أن النوعية كقاعدة لفاعلية الاسترجاع.

وإذا كان البعض يفترض أن نوعية التكشيف تعنى "فاعلية الاسترجاع" الخاصة بالتكشيف، فهناك من لا يرى هذا الاتجاه .. فالباحث رولنج (Roling, 1981) على سبيل المثال يرى أنه يمكن تعريف التكشيف بأنه درجة الاتفاق بين المصطلحات المخصصة بواسطة المكشف ومجموعة المصطلحات المثالية .. وأن هذه المثالية تتحقق من خلال شكل ما من أشكال اتفاق الخبراء .. وأن العمل الذى يقوم به أحد المكشفين يمكن مقارنته بالاتفاق العام ..

خامساً: نوعية التكشيف وخاصية الانتظام :

قام لانكستر (Lancaster, 1991, 62-) بتحديد بعض هذه العوامل كما يلى:

أ – عدد المصطلحات التى يتم تعيينها.
ب- المصطلحات المحكومة بالمقارنة بتكشيف النص الحر.
ج – حجم ودرجة تخصيص المصطلحات.
د – خصائص المجال الموضوعى ومصطلحاته.
هـ- عوامل تعود للمكشف.
و – الأدوات المتاحة للمكشف.
ز – طول المادة التى يتم تكشيفها.

ويمكن مناقشة هذه العوامل باختصار.

أ – عدد المصطلحات التى يتم تعيينها:

إذا ما طلبت إلى عدد من المكشفين تعيين مصطلحات تكشيفية تبعاً لأهميتها بالنسبة لموضوع معين فى وثيقة معينة، فيفترض أن هناك درجة جيدة من

الاتفاق بالنسبة للمصطلحات على رأس القائمة، وكلما نزلت فى القائمة إلى أسفل سيقل عدد المصطلحات التى يتم الاتفاق عليها بين المكشفين .. أى أن هناك عادة اتفاقا بالنسبة للموضوعات الرئيسية للوثيقة وليس الأمر كذلك بالنسبة للموضوعات الفرعية.

ب- المصطلحات المحكومة بالمقارنة بتكشيف النص الحر:

إحدى المميزات الرئيسية للمصطلحات المحكومة هى أنها ستحسن عملية الانتظام فى تمثيل المادة الموضوعية .. ولكن هذه العلاقة ليست مباشرة كما قد يتبادر إلى الذهن .. فالمصطلحات المحكومة ستحسن الانتظام فى التكشيف على المدى الطويل، ولكن ذلك يجب أن يكون على يد مكشفين ذوى خبرة فى المجال الموضوعى فضلاً عن إحاطتهم الكاملة بالمصطلحات.

ويثير فوجمان (Fugman, 1985) قضية هامة بالنسبة للانتظام، فإذا كانت دراسات الانتظام تركز على اختيار المصطلحات لوثيقة معينة، فإن الباحث عن المعلومات يهتم أساساً بالانتظام بين الوثائق المتصلة بدراسته .. أى أن هناك مقياساً آخر لتحليل الانتظام يمكن أن يكون مفيداً .. أى ما هو المدى الذى يمكن أن يتم فيه التكشيف المنتظم خلال قاعدة البيانات.

وأخيراً فيمكن الإشارة لتجربة ماكين (Mc Cain et. al, 1987) حيث قام ببحوث إنتاج فكرى مقارن بالنسبة لعدد (١١) طلباً وضعها متخصصون فى العلوم السلوكية الطبية وتمت هذه البحوث المقارنة فى الميدلاين Medline والمستخلصات الطبية Excerpta Medica ومعلومات علم النفس Psyc Info وبحوث العلوم SCISEARCH وبحوث العلوم الاجتماعية Social Sciscarch ..

وقد تمت البحوث فى قواعد المعلومات الثلاثة الأولى على:

أ- المصطلحات المحكومة. ب- اللغة الطبيعية.

كما تم فى قواعد بيانات الاستشهادات المرجعية على ما يلى: أ - استخدام اللغة الطبيعية للعناوين. ب - استخدام الاستشهادات لمواد مفيدة معروفة كنقاط مداخل، وعلى الرغم من أن الغرض الأساسى هو دراسة نوعية التكشيف فى

ميدلاين Medline .. إلا أن القليل من النتائج أو التوصيات قد تم الوصول إليها ومن بينها:

١ - أن إدخال اللغة الطبيعية فى استراتيجية البحث أدت إلى تحسين ملحوظ فى الاستدعاء بالمقارنة باستخدام المصطلحات المحكومة وحدها ٢- ليس هناك قاعدة بيانات يمكن أن تزودنا بمفردها بالتغطية الكاملة لإنتاج فكرى معقد ومتعدد الارتباطات.

ج - حجم ودرجة تخصيص المصطلحات:

كلما زاد عدد المصطلحات زادت درجة التخصيص، وكلما زاد التخصيص أصبح الانتظام أكثر عسراً وصعوبة (Tinker, 1968) .. فعلى سبيل المثال يمكن لاثنين من المكشفين الاتفاق على مصطلح "تآكل المعادن" Corrosion كمصطلح كشفى لوثيقة معينة، ولكنه من غير المحتمل اتفاقهما على أنواع من التآكل تتم مناقشته.

د - خصائص المجال الموضوعى ومصطلحاته :

يفترض أن انتظاما أكبر سيحدث فى تكشيف الموضوعات الأكثر صلابة (مثل الأشياء المادية أو أسماء الأفراد) .. وأن الانتظام سيقل عند معالجة الفرد للتجريدات.

هـ - عوامل تعود للمكشف :

يمكن أن نتوقع أن اتفاق اثنين أو أكثر من المكشفين فى التعليم والخبرة والاهتمامات، سيؤدى إلى أن يكونوا أكثر اتفاقاً فى التكشيف من اثنين يختلفان تماماً فى خلفياتهم التعليمية والخبرات، وإن كان العالم. آلن كنت(*) يرى أنه من المستحيل على المكشف الفرد أو مجموعة المكشفين تحقيق الانتظام المطلق فى التحليل الموضوعى للوثائق، ولعل ذلك يعود فى رأيه للطبيعة الإنسانية حيث لا

(*) كنت، آلن (١٩٧٣) ثورة المعلومات: استخدام الحاسبات الإلكترونية فى اختزان المعلومات واسترجاعها، ترجمة حشمت قاسم، وشوقى سالم ومراجعة أحمد بدر. الكويت: وكالة المطبوعات.

يحتمل قيام الإنسان بالاختيار نفسه على طول الخط حتى وإن كانت ظروف الاختيار واحدة من حيث عناصرها المؤثرة (كنت، آلن ١٩٧٣).

و – الأدوات المتاحة للمكشف:

إذا كانت جماعة المكشفين تشترك معاً فى مجموعة مشتركة من الأدوات المساعدة (كالقواميس وكتب الحقائق ...) فإن هذه الجماعة تميل إلى تحسين الانتظام بين هذه الجماعة.

ز – طول المادة المكشفة :

طول المادة يؤثر على مدى الانتظام فى التكشيف، وكلما قصرت المادة، قل عدد المصطلحات التى يمكن استخدامها.

سادساً : بعض مشكلات التكشيف واقتراحات الحلول

هناك مشكلات عديدة فى التكشيف يرصدها الباحثون والممارسون، منها أن التكشيف لا يستجيب لاحتياجات الباحث، لأن التكشيف يتناول بطريقة عامة فقط ماذا تقول الوثيقة aboutness، كما أن التكشيف لا يركز على ما تقدمه الوثيقة من جديد يتصل بموضوع البحث .. ويفسر بعض الباحثين مثل وايبنبرج (Weinberg, 1988) ذلك بأن هناك فرقاً بين وجهة النظر Aspect أو أحد جوانب الموضوع وبين ما تتحدث عنه الوثيقة أو تقوله Aboutness كما أن هناك فرقاً بين الموضوع Topic وبين التعليق Comment ..

ومع ذلك فيرى باحثون آخرون. أن ما تقوله الوثيقة Aboutness مفهوم غامض، لأن المحتوى الموضوعى للوثيقة يمكن أن يرجع إليه على أنه "قول داخلى intrinsic aboutness ، أما ما يمكن أن تستخدم فيه الوثيقة أو سبب اقتنائها أو غير ذلك من الاعتبارات والمتغيرات الخارجية فيرجع إليه على أنه قول خارجى extrinsic aboutness

وهناك مشكلة من الباحثين من يرون فى الذكاء الاصطناعى وسيلة لحل مشكلات اختيار المصطلحات الصالحة للتكشيف من أجل استرجاع أفضل، أى الاستعانة بالنظم الخبيرة التى تصمم لمعاونة المستفيد فى صياغة أسئلته واختيار الوثائق المتصلة (Croft, 1987).

وهناك مشكلة المجالات المعرفية غير اللفظية Nonverbal مثل الخرائط الجغرافية أو الجيولوجية وهى مجالات بصرية Visual بالدرجة الأولى، وبالتالى فلا تتم خدمتها بدرجة كافية بالنظم اللفظية Verbal وهناك من يقترح نظام الإيديوجراف Ideograph كنموذج لتصميم قاعدة البيانات والتكشيف واسترجاع المعلومات الجيولوجية (Grande, 1987) أما الباحث يتس (Yeastel, 1988) فيتناول متطلبات التكشيف فى الفيديوتكس والنظم القريبة مع الإشارة إلى نظام بريستل Prestel البريطانى. حيث تتناول مداخل تكشيف البريستل ما يلى:

- الأطر المرجعية للقائمة المحكومة Menu-Controlled reference frames
- الكشافات الهجائية لكل من الموضوعات وأسماء الأشخاص.
- كشافات الكلمات المفتاحية.
- كشافات البوابات للنظم الأخرى Gateway indexing .
- الكشافات المطبوعة وعلى الخط المباشر.
- كشافات المستفيدين المحكومة User- controlled indexes .
- أرقام الصفحات التذكرية mnemonic page numbers .
- مؤشرات الصفحات Pagemarker وهذه تتيح للمستفيدين للتأشير على صفحات محدودة لاستدعائها السريع فيما بعد.



الفصل السادس

المستخلصات والاستخلاص
دراسة مسحية تحليلية

تعريف وتقديم:

يعرف المستخلص بأنه التمثيل الدقيق والموجز لمحتويات الوثيقة، وذلك بأسلوب شبيه بذلك الأسلوب الخاص بالوثيقة الأصلية، ودون أى إضافة أو تفسير أو نقد، وإن كان هذا التعريف قد تطور ليلائم الأشكال العديدة من المستخلصات وأساليب الاستخلاص فى مختلف الظروف والوثائق والمستفيدين.

كما يعتبر الاستخلاص جزءاً من التحليل الموضوعى، شأنه فى ذلك شأن التصنيف والتكشيف، لأنه إجراء يتم بقصد تمثيل المحتوى المعرفى للتسجيلات حتى يستطيع المستفيدون العثور على المعلومات التى يحتاجونها، ولكن الاستخلاص يختلف عن التصنيف والتكشيف من حيث تقديمه لبعض المعلومات الفعلية التى تحتويها الوثيقة، أى أن المستخلص يلخص المحتويات الأساسية للتسجيلة المعرفية، ويعتبر بذلك كبديل حقيقى للوثيقة نفسها فى بعض أنواع المستخلصات، ونظراً لأن دوريات المستخلصات تحتوى عادة على كشافات، فإن التكشيف يعتبر جزءاً مباشراً من عملية الاستخلاص (Cleveland, D., 1990, p. 160) .

ولعل تاريخ الاستخلاص الحديث يعود إلى عام ١٦٦٥م عندما ظهرت مجلة العلماء Le Journal des Scavans فى باريس واحتوت على بعض المستخلصات النقدية للكتب وغيرها من وسائل الاتصال الأخرى، وإن كانت الحضارات القديمة قد شهدت فى مصر لفافات البردى ومعها مستخلصات لمحتوياتها. (Cleveland, D., 1990, p. 185) .

وشهدت حضارة بابل وآشور الكتابة على الألواح الطينية، حيث كانت الوثيقة تستخلص على الغلاف الخارجى (Borko, H., 1975, p. 26) ولكن المستخلصات فى الوقت الحاضر تلعب دوراً أساسياً فى بناء نظم المعلومات المحسبة، وأخيراً فإن المستخلصات ذات أهمية بالغة للمكشفين، ذلك لأن المستخلص المكتوب جيداً يحمل كلمات ومصطلحات تعتبر مصدراً قيماً للمكشف.

وإذا كان الإنتاج الفكرى المعلوماتى العربى يشير عام ١٩٦٤ إلى مقال أحمد بدر عن المستخلصات التقليدية والتلغرافية، واستخدام الأخيرة فى نظم المعلومات الآلية فى بداية الستينيات، فقد استوعب حشمت قاسم عام ١٩٨٤م التطورات التى حدثت خلال العشرين عاماً التالية (بما فى ذلك ما جاء عن المستخلصات والاستخلاص فى كتاب المكتبات المتخصصة عام ١٩٧١م الذى ألفه بالاشتراك مع أحمد بدر وما جاء فى كتاب ثورة المعلومات الذى ترجمه وزميله شوقى سالم وراجعه أحمد بدر) وإذا كانت يسرية زايد عام ١٩٩٤م قد أضافت بُعداً جديداً عن المستخلصات بالتركيز على المعايير الموحدة فستحاول الباحثة استكمال هذا الجهد العربى الذى استمر أكثر من ثلاثين عاماً، مستعينة فى دراستها هذه بالكتب الإنجليزية الحديثة عن الاستخلاص، فضلاً عن قيامها ببحث إنتاج فكرى فى قاعدة بيانات مستخلصات علم المعلومات (ISA) ولكن بالتركيز على التسعينيات.

هذا وتتناول الدراسة الحالية النقاط الثمان التالية:

أولاً : وظائف المستخلصات وأغراضها ومحتواها العام.

ثانياً : بعض مصطلحات المستخلصات وطبيعتها.

ثالثاً : اتجاهات البحوث والدراسات عن المستخلصات فى التسعينيات.

خامساً : من الذى يعد المستخلصات؟

سادساً : خطوات وإجراءات القيام بالاستخلاص.

سابعاً : كيفية كتابة وتحرير المستخلص.

ثامناً : نماذج لأنواع المستخلصات الرئيسية ولخدمات الاستخلاص فى مخلف المجالات العلمية.

أولاً : وظائف المستخلصات وأغراضها ومحتواها العام :

أ – وظائف وأغراض المستخلصات:

تعد المستخلصات لمعاونة الباحث فى تقييم محتويات الوثيقة وأهميتها المحتملة بالنسبة له. واكتساب مهارات إعداد المستخلصات ليست مفيدة لكل من الأمين وإخصائى المعلومات فحسب بل للطالب والباحث كذلك، فهى تساعد الأخير على أخذ الملاحظات الهامة واستيعاب الإنتاج الفكرى الجارى وتحليل البحوث وتقديم التقارير.

كما أن المستخلصات ذات أهمية قصوى بالنسبة لعمليات اختيار الوثائق وتجميع المعلومات والمساعدة فى تجنب التكرار والتأخر فى الإعلام عن البحوث الجارية وبالتالى فالمستخلصات موجودة فى معظم المطبوعات الأولية والثانوية.

والمستخلصات فى المطبوعات الأولية تعد عادة بواسطة المؤلفين أنفسهم وهؤلاء لم يتدربوا على أن يكونوا مستخلصين. كما أن المستخلصات هى العمود المحورى للمطبوعات الثانوية، والمستخلصات بالتكاتف مع الكشافات تشكل قطاعاً أساسياً فى خدمات النشر الثانوى، وتستخدم مراصد البيانات المحسبة فى الوقت الحاضر هذه المستخلصات كأداتها الرئيسية فى تمثيل الوثائق. والعديد من خدمات الإحاطة الجارية والراجعة والمعتمدة على مراصد المعلومات هذه تحتوى على المستخلصات وإن كان إعداد المستخلصات وإدخالها فى خدمات الإحاطة الجارية يعطل طبيعتها الجارية نظراً لما يتطلبه هذا الإعداد من وقت وإن كانت الحاسبات الآلية قد يسرت هذه العملية إلى حد كبير. أما بالنسبة للبحث الراجع فمن أبعد الأمور بل من استحالتها فرز الأعداد الهائلة من الوثائق الكاملة فضلاً عن صعوبة وجودها فى مكتبة واحدة وبالتالى تصبح مستخلصات تلك الوثائق فى البحث الراجع ذا أهمية بالغة.

ويتضح مما سبق أن المستخلص يتم إعداده بغرض توفير وقت المستفيد بالنسبة لتجميع المعلومات واختيار المتصل منها باهتماماته وبالتالى يستطيع المستفيد اتخاذ القرارات السليمة بالنسبة للبحث والتطوير وتقليل كمية التكرار فى البحوث لاسيما مع فيضان المعلومات المعاصر وعدم استطاعة أى باحث متابعة التطورات والمكتشفات فى مجاله المتخصص. وهذا الأمر ليس قاصراً على الباحثين بل ينسحب أيضاً على المديرين والأساتذة والمعلمين والذين يحتاجون إلى متابعة كل جديد فى مجالات أوسع نسبياً من الباحثين وبطريقة عامة أيضاً.

هذا وتساعد المستخلصات على تخطى الحواجز اللغوية، ذلك لأنها تمكن المستفيد من الحكم على ضرورة عمل الترجمة للمقال أو البحث أو استبعاد ذلك ولعل هذه المستخلصات ــ بالذات الإعلامية ــ تعفيه من ترجمة المقال الأصلى وأخيراً فتساعد هذه المستخلصات فى تجميع وإعداد الأدوات الببليوجرافية الأخرى كالكشافات والببليوجرافيات والمراجعات (Rowley, J., 1982).

ويمكن التركيز على وظائف المستخلصات فيما يلى :

(١) تيسير وتسهيل عمليات الاختيار، أى أنها تساعد القارئ على اتخاذ قراره بشأن مادة أو وثيقة معينة ودرجة ارتباطها بتخصصه ودراساته.

(٢) توفر وقت القارئ وفى حالة المستخلص الإعلامى الجيد فسيكون فى الواقع بديلاً لقراءة المادة الأصلية نفسها.

(٣) التعرف على محتويات مواد معينة مكتوبة بلغات غير مألوفة لقارئ معين.

(٤) متابعة الباحثين والقراء للإنتاج الفكرى الجديد المنشور فى مجالاتهم (كما هو الحال فى الإحاطة الجارية Current Awareness والبث الانتقائى للمعلومات (SDI).

Selective Dissemination of Information.

(٥) يمكن للمستخلصات أن تلعب دوراً هاماً فى نظم الاسترجاع المحسبة، وذلك بتيسير التعرف على المواد ذات العلاقة والارتباط والصلاحية فضلاً عن تزويدنا بنقاط الإتاحة للمواد المختزنة (وذلك فى النظم التى تختزن نصوص المستخلصات فى شكل يصلح للبحث).

(ب) **المحتوى والشكل العام للمستخلص** Content & Format

ما ينبغى أن يشمله المستخلص يعتمد من غير شك على نوع المطبوع الذى بأيدينا، فالمستخلص الشارح الطويل لبعض أنواع تقارير البحوث قد يتضمن أهداف البحث والإجراءات والتجارب المستخدمة والنتائج التى تم تحقيقها (المستخلص الإعلامى سيقدم لنا النتائج الفعلية على الأقل فى شكل مكثف)، فضلاً عما انتهى إليه المؤلف من بيان عن دلالة النتائج، أما معالجة المقال التاريخى فهو أمر مختلف تماماً، فقد يتضمن المستخلص على سبيل المثال تأكيداً لرسالة المؤلف أو ما انتهى إليه، مع الاهتمام بذكر الفترات والمناطق الجغرافية والأشخاص الذين ذكرهم بدراسته.

أما بالنسبة للمجالات الموضوعية المتخصصة، فقد يُتطلب من المستخلص اتباع تعليمات محددة بالنسبة لأشياء معينة خاصة بالنسبة للبحث فى المقال عن أشياء معينة

وإبرازها، كما هو الحال عند بيان جرعات دواء معين أو أحوال مناخية أو عمر أفراد معينين أو أنواع التربة أو المعادلات المستخدمة .. إلخ، وتصبح المستخلصات أكثر يسراً فى كتابتها عندما يتناول الموضوع أشياء محسوسة، كما تصبح كتابة المستخلص أكثر صعوبة عندما يكون الموضوع مجرداً.

هذا وتقدم معظم المستخلصات فى الشكل التقليدى، حيث توجد المداخل الببليوجرافية يتلوها نص المستخلص وفى بعض المطبوعات فإن المستخلص يسبق المداخل الببليوجرافية، وقد يكتب السطر الأول بالحروف الكبيرة، وهذه قد تجذب انتباه القارئ شأنها شأن العناوين الصحفية، ويرى البعض أن عنوان المقال يمكن أن يكون هو السطر الأول البارز، وخلاصة ما سبق فالمستخلص الكامل يحتوى على ثلاثة أجزاء :

الأول: هو المصادر المرجعية الببليوجرافية للمادة المستخلصة.

الثانى : نص المستخلص.

الثالث: التوقيع سواء بالاسم نفسه أو بالحروف الاستهلالية، حيث يدل ذلك على أن القائم بالاستخلاص شخص آخر غير المؤلف، أو المؤلف نفسه (Lancaster, 1991, p. 101).

ومعظم المستخلصات تقع فى حوالى ١٠٠-٢٥٠ كلمة وإن كانت هناك عوامل أخرى تتحكم فى طول المستخلص مثل طول الوثيقة نفسها ودرجة تعقد المحتوى الموضوعى، ومدى أهميتها، ومكانية الوصول المادى لها (أوراق المؤتمرات أو أعمال مكتوبة بلغات نادرة ... إلخ).

وقد أوصى العالمان بوركو وبيرنير (Borko, H., 1975) أن تحتوى مستخلصات الإنتاج الفكرى العلمى ما بين $\frac{١}{١٠}$ ، $\frac{١}{٢٠}$ من طول الوثيقة نفسها الأصلية، وقد قام الباحثان بتقديم بعض النصائح عن تتابع المحتوى كما يلى:

يمكن أن يرتب المستخلص بطريقة توفر وقت القارئ، حيث توضع النتائج فى البداية، وبالتالى توفر على القارئ المضى فى القراءة، فقد يقوم القارئ بتقبل هذه النتائج أو رفضها دون الحاجة إلى التعرف على أدلة هذه النتائج، كما أنه لا داعى لوضع عناوين فرعية مثل كلمة نتائج أو إجراءات أو مناهج .. إلخ، ذلك لأن القارئ

سيفهم ذلك دون ذكره، كما أنه لا داعى لتقسيم المستخلص إلى فقرات، فالمستخلص نفسه قصير، ويجب أن يعبر عن التجانس فى الفكر، أى أن المستخلص يمكن أن يكتب كفقرة واحدة.

ثانياً : بعض مصطلحات المستخلصات وطبيعتها :

قبل التعرف على الأنواع الرئيسية للمستخلصات والتمييز بينها، لابد فى البداية من الإشارة إلى أن هناك مصطلحات عديدة باللغة الإنجليزية وليس لها نظير دقيق فى معناها باللغة العربية، وستحاول الباحثة الإشارة فى هذا الصدد إلى المصطلحات التالية: الحاشية Annotation/ الاقتباس Extract/ الملخص Summary/ التصغير Digest-Paraphrase- Precis Abridgement/ المستخلص المصغر Mini- Abstract المستخلصات التلغرافية – المستخلصات القياسية.

وقد جمعت الباحثة هذه المصطلحات من كتب عديدة فى الاستخلاص أهمها كتب لانكستر وبوركو وراولى وكليفيلند.

(١) فالشرح أو الحاشية Annotation

تعنى مذكرة مضافة إلى العنوان والبيانات البيليوجرافية للوثيقة عن طريق الشرح لمحتوياتها وفى الفهرسة فإن هذه الحاشية أو المستخلص الشارح يستخدم على البيانات التى ترفق بالوصف الوراقى المعيارى للوثيقة وكجزء من مدخل هذه الوثيقة فى الفهرس. وقد يحتوى هذا الشرح على تعليقات على أى عنصر من عناصر الوصف الوراقى. وبالتالى فالحاشية فى الوراقية هى أقرب ما تكون إلى المستخلصات وإن كانت أهدافها أكثر، محدودية من المستخلص وأكثر اختصاراً منها.

(٢) أما المقتبس Extract

فهو شكل آخر من أشكال التعبير عن الوثيقة. وهو يشمل جزءاً أو أكثر من الوثيقة وهذا الجزء يختار لتمثيل الوثيقة كلها، ولكن لابد أن تؤخذ كلمات المقتبس من نفس النص وعادة من الأجزاء الخاصة بالنتائج والتوصيات، وعلى ذلك فطبيعة المقتبس هذه لا تجعله يمثل الوثيقة بطريقة متوازنة وإن كان سيبرز بعض النقاط الدالة الموجودة بالوثيقة.

ومن هذا المنطلق **فيعرف الاقتباس:** بأنه صيغة مختصرة للوثيقة تم إنشاؤها عن طريق اختيار بعض الجمل من الوثيقة نفسها، وذلك باختيار جملتين أوثلاثاً من المقدمة، يتبعها جملتين أو ثلاثاً من النتائج أو الملخص، حيث يمكن أن توضح هذه الجمل المختارة شرحاً جيداً لمقال الدورية مثلاً وماذا تعنيه.

(٣) أما **الملخص** Summary

فلا يستطيع المدقق التمييز بينه وبين المستخلص، وإن كان الملخص على وجه التحديد هو إعادة صياغة الوثيقة باختصار أى أنه ملخص الوثيقة بترتيب أجزائها المختلفة وبالتالى فهو يقدم للقارئ المعالم البارزة للوثيقة ويفترض هنا أن القارئ سيرجع لهذه الوثيقة بعد ذلك على عكس المستخلص (الإعلامى) الذى يفترض ألا يعود القارئ بعد قراءته إلى الوثيقة الأصلية، أى أن المستخلص يعفيه من الرجوع للأصل ولكن الملخص لا يعفيه من ذلك.

(٤) مصطلحات أخرى:

هناك مصطلح ABRIDGMENT وهو Reduction تصغير للوثيقة الأصلية وبالتالى تحذف عدد من النقاط الثانوية لأنه مصطلح عام نسبياً، أما مصطلح Precis فهو يعنى بياناً عن الوثيقة يحصر نفسه فى النقاط الضرورية جداً للمناقشة.

أما مصطلح PARAPHRASE فإنه يعنى تفسيراً للأفكار المسجلة بالوثيقة وترجمته للغة الكاتب الذى أعد هذا الموجز. أما مصطلح DIGEST فهو ترتيب منهجى لتقديم المناقشات الرئيسية فى الوثيقة، وأخيراً فإن نبذة المؤلف Synopsis هى المصطلح المستخدم أصلاً للدلالة على الموجز المعد بواسطة مؤلف العمل العلمى وواضح أن التمييز بين هذه المصطلحات أمر عسير للغاية، كما أن هذه المصطلحات لا تستخدم بمعنى واحد بشكل منتظم وعلى ذلك **فالمستخلصات القصيرة** Short Abstracts والتى تحتوى على جملة أو جملتين فقط يمكن أن تكون كافية فى خدمات المعلومات التجارية. وكذلك فإن الكلمات المفتاحية أو المصطلحات الكشفية والتى تأتى مع الاستشهاد المرجعى يمكن أن تخدم ككشاف عام عن المجال الموضوعى. وقد تناول لانكستر **مصطلح المستخلصات المصغرة** Mini Abstracts على أنه مصطلح غير دقيق ذلك لأنه يمكن أن يعنى مستخلص مقيد، ومع ذلك فقد

استخدمه العالم (Lunin, 1967) ليدل على المستخلص عالى التركيب، والمصمم أساساً لتيسير البحث بالحاسب الآلى، وهو فى واقع الأمر نوع من التزاوج بين المستخلص والمدخل الكشفى، وقد عرّفه Lunin بأنه المستخلص الكشفى القابل للقراءة بالآلة، ويحتوى المستخلص المصغر على مصطلحات محكومة، ولكنها توضع فى تتابع محدد، وعلى سبيل المثال فباللغة الإنجليزية يمكن أن تسجل الجملة التالية، ثم المستخلص المصغر الدال عليها: الجملة There is adecreased amount of zinc in the blood of humans with cirrhosis of the liver. وترجمة الجملة هو أن هناك كمية متناقصة من عنصر الزنك فى دم الإنسان المصاب بتليف فى الكبد المستخلص الصغير./ **Decr/Zinc/Blood/Humans/Cirrhosis/ liver.**

ويلاحظ هنا أن القائم بعملية الاستخلاص يحاول الحفاظ على تتابع المصطلحات على قدر الإمكان بحيث تتفق مع تركيب الجملة العادى، كما أن محتويات الوثيقة يمكن وصفها بشيء من التفصيل من خلال استخدام سلسلة الجمل ذات الأسلوب السابق، وعلى الرغم من أن هذه المستخلصات قد أعدت أساساً لتيسير البحث بالحاسب الآلى إلا أن هذه المستخلصات المصغرة ذات دلالة للقارئ الذكى.

وهناك ما يسمى **بمستخلصات تركيز الأضواء** Highlight Abstracts ويصمم هذا النوع لجذب القارئ نحو مقال معين وفتح شهيته لقرائته ومستخلصات تركيز الأضواء هذه تظهر فى العديد من الدوريات الأولية كمقاطع تسبق المقال المعلقة به أو كملحق بعد قائمة المحتويات. وليس هناك ادعاء بأن هذه المستخلصات متوازنة أو تمثل صورة كاملة للمقال، وفى الواقع فهى تثير الاهتمام لدى القارئ ولعل هذا فى حد ذاته بداية الإفادة الحقيقة من المقال، ومعنى ذلك أن مستخلصات تركيز الأضواء هذه لا تستطيع أن تعيش أو تنتشر بمفردها دون وجودها مع المقال نفسه.

أما **المستخلصات الإحصائية والجدولية والعددية** فهذه تعتبر وسيلة لتلخيص البيانات الرقمية، والمهارة مطلوبة هنا لاختيار البيانات ذات الدلالة من الجداول الأصلية أو التمثيل الخطى والرسومات ثم وضعها فى إحصائيات حيوية وبشكل مكثف. وتتضمن هذه المستخلصات من غير شك أشكالاً من البيانات الاقتصادية والاجتماعية والتسويقية.

(Rowley, 1992, pp. 9-16)

(٥) المستخلصات التلغرافية Telegraphic Abstracts

مصطلح المستخلص التلغرافى هو مصطلح غير دقيق ذلك لأنه يعنى تمثيل الوثيقة وتقديمها بطريقة اقتصادية مختصرة، أى أنه يشمل الجملة الكاملة، ولكنه يشمل ما يشبه الكلمات التلغرافية، ولعله يدل على خيط من المصطلحات دون أن يكون بينها رابط نحوى. وعلى ذلك فمصطلح المستخلص التلغرافى قد استخدم للدلالة على المكونات الأساسية لنظام الاسترجاع المحسب، والذى تم تطويره فى جامعة Case Western Reserve بأمريكا.

هذا والمستخلصات التلغرافية عبارة عن تحليل موضوعى للوثيقة فى شكل يصلح "للبحث الآلى". والواقع أن الخلاصة التلغرافية ما هى إلا كشاف تحليلى وتفصيلى للمقال أى أنها تحتوى على قائمة كلمات دالة Significant من المقالة الأصلية، وهذه الكلمات منظمة فيها بينها طبقاً للتركيب اللغوى Syntactics الذى يبين علاقة كل كلمة بالأخرى (التى تسبقها أو التى تليها) وذلك عن طريق:

(أ) الدلالات (Role indicators)

(ب) رموز الترقيم (Punctuations)

ويتم عمل بطاقات مثقوبة للكلمات المستخدمة فى الخلاصة التلغرافية (Key Punched Cards) ثم تقابل (matched) على قاموس المعانى للكلمات الإنجليزية English Word Semantic Code Dictionary file حيث يتم تكويد الكلمات آلياً.

ومعنى ذلك أننا نستبدل الكلمة الإنجليزية المكتوبة بها المقالة العلمية إلى كلمات (أورموز) لا يقرأها الإنسان وإنما تقرأها الآلة فقط، ويعبر عنها بلغة الآلة Machine Language وهى كلمات مركبة تدل على جذور المعانى الخاصة بالكلمة الأصلية، ومعنى ذلك أن الكاتب إذا استخدم مثلاً كلمتين لها نفس المعنى، فيكون للكلمتين كود للمعانى واحد فقط وكود المعانى هذا (Semantic Code) يخدم الأغراض التالية:

(١) يعمل على التحكم فى المصطلحات المستخدمة فى المقالات المختلفة الموجودة فى الخلاصات التلغرافية.

(٢) يحول اللغة الطبيعية التى يقرأها الإنسان إلى لغة الآلة أو لغة الحاسب الآلى.

ومع التطور الهائل لاختزان المعلومات واسترجاعها بواسطة الحاسبات الآلية، وكذلك التطور الملحوظ فى المكانز، أصبح نظام جامعة كيس وسترن ريزرف على الرغم من أهميته الأساسية فى علم التوثيق والمعلومات، يعتبر فى الوقت الحاضر ذا أهمية تاريخية فحسب (أحمد بدر، ١٩٦٤) هذا ويشير لانكستر إلى مصطلح المستخلصات القياسية .Modular Abst والمقصود بها الوصف الكامل لمحتوى الوثائق الجارية، حتى يحتوى كل مستخلص على الأجزاء الخمسة التالية:

(١) الاستشهاد المرجعى Citation.

(٢) الحاشية Annotation.

(٣) المستخلص الشارح .Indicative Abstract.

(٤) المستخلص الإعلامى .Informative Abstract .

(٥) المستخلص النقدى .Critical Abstract.

وتصمم هذه المجموعة بحيث يمكن تجهيز خدمة الاستخلاص لتتلاءم مع الاحتياجات الفعلية بأقل المجهود. والغرض الـرئيسى للمستخلصات القياسية هو التخلص من التكرار وضياع الجهد الفردىً الفكرى فضلاً عن تقليل الأعمال الفرديـــة التى تتم بالنسبة لنفس الوثائق.

ويلاحظ أن المداخل الكشفية بالمستخلصات يقوم بإعدادها متخصصون موضوعيون،

نوع الرى	نوع التربة	المحاصيل	الأحوال الجوية	المكان	النتائج

والأشكال القياسية ستؤدى إلى تقليل المعالجة التكرارية وسرعة تدفق العمل داخل خدمات الاستخلاص (Lancaster, 1991, p. 96) كما يضيف لانكستر أن هناك بعض. المواقف أو الأغراض التى يفضل فيها المستخلص التركيبى على المستخلص فى الشكل الروائى أو السردى Narrative ولعل الإطار أعلاه للمستخلص البنائى Stuctured Abstract فى موضوع الرى يوضح لنا هذه الفكرة (Lancaster, 1991, p. 91)

وفى المثال السابق فالمجال الموضوعى هو الرى؛ وبالتالى فالقائم بعملية الاستخلاص سيبحث عن وصفة محددة من المواد الموجودة فى الإطار (الجدول)، ويتضمن الاستخلاص فى هذه الحالة وضع القيم المناسبة داخل الإطار، وقد تستخدم الأكواد للتعبير عن النتائج التى تم الوصول إليها، وهذا النوع من المستخلصات له أهمية فى إعداد كتب الحقائق التى تلخص عدداً هائلاً من الدراسات التى تتم فى حقل معين، ومع ذلك فهذه الطريقة تصلح عادة عند ثبات العناصر الرئيسية فى المجال الموضوعى بالنسبة للدراسات المختلفة، كما أوضح الباحث سولكوفا (Zholkova, 1975) كيفية استخدام التحليل الوجهى لإنشاء المستخلص التركيبى، وإن كان هذا الاتجاه لم يأخذ نصيبه الكافى فى التطبيق العملى فى الإنتاج الفكرى.

وأخيراً فقد تصنف المستخلصات حسب الغرض الخارجى أى حسب كونها **مستخلص لموضوع محدود** discipline oriented أو حسب كونها كتبت لدعم أنشطة تطبيقية قد تكون متداخلة موضوعياً فيما بينها أولاً تكون وتسمى فى هذه الحالة Mission oriented أى موجهة لتحقيق رسالة معينة . (Cleveland, 1990 166) وهناك المستخلص الذى يركز على جزء مختار من المحتوى الموضوعى للوثيقة ويسمى Slanted abstract وهو يعتبر جزءاً من مستخلصات تحقيق الرسالة mission oriented (كما هو الحال مثلاً عند مناقشة أمراض الكلاب والقطط والدواجن وفى حالتنا هذه نناقش فقط أمراض الكلاب لاهتمام فريق البحث بهذا الجانب فقط) (Cleveland, 1990, p 166).

أى أنه ينبغى التمييز بين الخدمات التى تدور حول موضوع علمى محدد وبين الخدمات التى تدور حول رسالة معينة Disipline oriented & Mission oriented ، والخدمات الأولى التى تتصل بالكيمياء أو علم الحياة أو العلوم الاجتماعية على سبيل المثال، أما الخدمات الموجهة لرسالة معينة فهى التى تستجيب لصناعة معينة أو جماعة من الأفراد، كما هو الحال فى المستخلصات التى تتم لصناعة المطاط، أو المستخلصات المعدة للعاملين فى مجال التمريض، وبالتالى فإن المستخلص الموجه Subject Slanting سيكون أكثر ارتباطاً بالخدمات الموجهة لموضوع محدد، ذلك لأن اهتمامات المستفيدين فى حالة التوجه

نحو رسالة معينة، هذه الاهتمامات تكون أكثر تجانساً وتخصصاً من اهتمامات المستفيدين الآخرين (Lancaster, F. W. 1991, 88).

ثالثاً: الأنواع الرئيسية للمستخلصات:

يعترف معهد المعايير الأمريكى بثلاثة أنواع من المستخلصات وهى (الشارحة indicative والإعلامية Informative والمزج بينهما) ولكن هناك من الباحثين من يرى وجود خمسة أنواع من المستخلصات (Borko, H. 1975) وهى الشارحة/ الإعلامية/ المزيج بينهما/ المستخلص النقدى ثم المستخلص الموجه لغرض متخصص).

ويقدم لنا الملحق (١) أمثلة من كل نوع من هذه المستخلصات وستحاول الباحثة شرح هذه الأنواع والتى جاءت فى معظم الكتب عن موضوع الاستخلاص إلى جانب بعض الأنواع الأخرى الموجودة فى الإنتاج الفكرى الحديث:

(١) المستخلصات الشارحة (الواصفة): Indicative

هى تلك المستخلصات التى تدل القارئ على ما سيجده إذا اطلع على المقال، ولكن المستخلص لايستنتج هذه المعلومات، فالمستخلصات الشارحة تحتوى عادة على كلمات كالتالى:

"تشكل/تناول/تناقش/ تصف.." وما يشبهها من كلمات، ولكن هذه المستخلصات نادراً ما تحتوى على نتائج فعلية، وبالتالى فهذه المستخلصات تساعد المستفيدين على الحكم على مدى صلاحية المقال الأصلى، فضلاً عن أنها تزودنا بنقاط إتاحة إضافية على الخط المباشر، ولكن هذا المستخلص الشارح لا يمكن أن يكون بديلاً عن المقال الأصلى.

وقد جاء فى دليل إريك (ERIC, 1982, VI) ما يلى:

المستخلص الشارح هو وصف لمحتويات وشكل الوثيقة أو مرشد لها، على أن يتم كتابة هذا المستخلص من وجهة نظر قارئ محاط علماً بالموضوع وإن يكون موضوعياً غير متحيز بالدرجة الأولى، هذا والمستخلص الشارح يدلنا بصفة عريضة ما تمت مناقشته أو شموله فى الوثيقة وبأى طريقة تم تقديم المعلومات، وفى حالة الضرورة إلى من توجه الوثيقة.

هذا وتفضل العديد من خدمات التكشيف والاستخلاص إنتاج المستخلصات الشارحة، نظراً لأنها أسهل وأسرع فى كتابتها من أنواع المستخلصات الأخرى، وفى الواقع فطبقاً لتعليمات ERIC فإن "المستخلص يمكن أن يحصل على معلومات كافية لكتابة المستخلص الشارح عن طريق فحص قائمة المحتويات والمقدمة والملخص ... إلخ أو عن طريق تصفح النص والتقاط بعض نقاط الإتاحة" والمستخلصات الشارحة قد تكون ضرورية لبعض أنواع الوثائق الأصلية ثم البيليوجرافيات وأعمال المؤتمرات والمقالات الاستعراضية وغيرها من الوثائق الأخرى التى يختلف محتواها بدرجة كبيرة هذا ويرى الباحث Cremmins أن المستخلصات الشارحة يمكن أن تحتوى على معلومات عن الغرض والنطاق أو المنهجية، ولكنها لا تحتوى على معلومات عن النتائج أو التوصيات، والمستخلص الإعلامى قد يحتوى أيضاً على معلومات عن الغرض والنطاق ومنهج البحث، ولكنه يجب أن يحتوى أيضاً على النتائج أو التوصيات، وقد يستخدم المستخلص الإعلامى كبديل معقول للوثيقة نفسها أى أننا فى هذه الحالة قد نستغنى عن قراءة الوثيقة، ولكن المستخلص الشارح لا يخدم عادة كبديل للوثيقة، أى أن المستخلص الشارح يبين للقارئ ما إذا كان يحتاج إلى قراءة الوثيقة الأصلية أم لا، كما أن المستخلص الشارح يكون أكثر صعوبة فى كتابته، وفى الواقع فمن الممكن كتابة المستخلص الإعلامى للدراسة التجريبية، ولكنه من الصعب والعسير بل من المُستحيل أحياناً القيام بنفس الشىء بالنسبة للدراسات النظرية والدراسة التى تتضمن الآراء. من أجل ذلك فإن المستخلصات الإعلامية موجودة بكثرة فى ممارسة العلوم والتكنولوجيا، أى أكثر من تلك التى توجد فى العلوم الاجتماعية أو الإنسانيات (Cremmins 1992).

ومما سبق فالمستخلص الواحد يمكن أن يتضمن عناصر من الجانب الشارح والجانب المعلوماتى، وذلك اعتماداً على اهتمامات القارئ المقصود، وعلى سبيل المثال فإذا تم عمل استخلاص لتقرير على تلوث الهواء، ونشر هذا المستخلص متوجهاً إلى الكيمائيين، فإن معظم الوارد فى المستخلص عن الجوانب البيئية ستكون عادة جانباً شارحاً، كما يتضمن جزءاً آخر من التقرير على مستخلص إعلامى يقدم النتائج بناء على مختلف التحاليل التى تمت على عينات من الجو، وعلى ذلك فالمستخلصات الشارحة أكثر شيوعاً وهى تصلح أساساً لتمثيل المناقشات والمقالات الاستعراضية والكتب وفى بعض الأحيان تصلح لبحوث

المؤتمرات والتقارير لا تحتوى على نتائج وكذلك المقالات القصيرة والببليوجرافيات.

(٢) المستخلصات الإعلامية Informative Abstracts

يعتبر هذا النوع من المستخلصات الشكل المفضل لدى معهد المعايير أنزى (ANSI) ذلك لأن المستخلصات الإعلامية يمكن أحياناً أن تحل محل الوثيقة الأصلية، لأنها تحتوى على النتائج الفعلية أو وجهات النظر الموجودة فى الأصل.

وطبقاً لدليل إريك (ERIC, 1982, VI-309) فيعتبر المستخلص الإعلامى صيغة مكثفة للأفكار الأساسية الموجودة فى الوثيقة وهو يحتوى على بيان من وجهة نظر الرسالة أو التطور أو الدليل أو النتائج، وباختصار فإن هذا المستخلص يقرر ماذا تقوله الوثيقة فعلاً.

هذا والكتابة الجيدة للمستخلص الإعلامى يمكن أن تكون أكثر صعوبة من المستخلص الشارح، ذلك لأن المستخلص الإعلامى يتطلب قراءة كل (أو معظم) الوثيقة الأصلية ثم تلخيص نقاطها الرئيسية بدقة، والمهم هنا أن هذا المستخلص يحتوى على وجهات نظر المؤلف الأصلى وليس أبداً صوت القائم بعملية الاستخلاص.

وهناك بعض المقالات التى تسمح معايير (ANSI) بدمج كل من المستخلص الشارح والمستخلص الإعلامى بالنسبة لها، وعادة يكون هذا النوع المزجى من المستخلصات صالحاً للوثائق الطويلة والتى لا يمكن تغطيتها بصورة كاملة بطريقة إعلامية، وهنا يتم استخلاص أكثر الأجزاء أهمية فى الأسلوب الإعلامى، بينما يقوم القائم بالاستخلاص بإعداد بيانات شارحة للجوانب أو الأجزاء الأخرى من الوثيقة.

والأنواع الأخرى من المستخلصات لا يعترف بها معهد المعايير (ANSI) لكنها تستخدم بواسطة خدمات الاستخلاص، فعلى سبيل المثال فإن المستخلص النقدى Critical Abstract يقدم لنا أحكاماً قيمية Value لمحتويات المقال، حيث يضمن القائم بالاستخلاص آراءه عن أهمية وجدارة المقال الذى يتم استخلاصه، ومعظم خدمات التكشيف والاستخلاص تحذر من هذه المستخلصات النقدية بل وينكر بعض الباحثين (Cleveland, D., 1990. 164) هذا النوع من المستخلصات،

ذلك لأنها لا يمكن أن تحمل كثيراً من المعلومات الأساسية وتعتبر فى واقع الأمر مراجعة review للوثيقة وليس بياناً حقيقياً عن محتواها" وأن النقد يجب أن يقتصر على المراجعات Reviews .

ويتضح مما سبق أن المستخلصات الإعلامية تقدم أكثر ما يمكن من المعلومات الكمية والنوعية التى تحتويها الوثيقة، وهذا المستخلص يستجيب لهدفين أولهما تقييم علاقة الوثيقة وارتباطها باهتمامات الباحث وبالتالى اختيارها محل الوثيقة الأصلية عندما تكون المعلومات الأساسية عن الوثيقة هى الكافية، فالمستخلص الإعلامى إذن يقدم لنا تركيزاً واضحاً للمناقشات الأساسية والنتائج التى وصل إليها الباحث فى الوثيقة.

ولتحقيق غرض المستخلص فهو يميل إلى أن يكون أكثر طولاً من المستخلصات الأخرى (من مائة على مائتين وخمسين كلمة وإن كانت الخمسمائة كلمة ربما تكون ملائمة للتقارير أو الرسائل العلمية) ومع ذلك فطول المستخلص يجب أن يتناسب مع حجم الوثيقة نفسها وبالتالى فليس هناك مقياس محدد لطول المستخلص الإعلامى وهو يعتبر مرغوباً أكثر من المستخلصات الكشفية Indicative وذلك للنصوص التى تشرح التجارب والوثائق التى تختص بمجال علمى محدد وتكون مكلفة فى تصويرها، ولكن هل يدل طول المستخلص عن جودته؟ يمكن مناقشة ذلك فيما يلى:

طول المستخلصات:

لا يدل طول المستخلص بالضرورة على جودته ونوعيته، وقد قررت ((ANSI أن المستخلص الذى يقل عن (٢٥٠) كلمة سيعتبر كافياً بالنسبة لمعظم المقالات أو أجزاء من الكتب، وقد أوصى كل من بوركو وبيرنير أن يكون طول المستخلص بين عشرة إلى واحد على عشرين من الأصل فى الإنتاج الفكرى للعلوم الطبيعية، وفى الواقع فإن الطول المرغوب للمستخلص يعتبر دالة لطول المقالات الأصلية وكذلك دالة لسياسة خدمة الاستخلاص.

ويقترح لانكستر ما يلى:

من المناسب أو المعقول أن يختلف طول المستخلص تبعاً لعدة عوامل منها طول المادة (المقال مثلاً) نفسها، مدى تغطيتها للموضوعات، أهميتها المتوقعة، إمكانية الحصول المادى على الوثيقة الأصلية فضلاً عن إمكانيات الإتاحة الفكرية (مثلاً مواد يصعب تحديد مكانها مثل أوراق المؤتمرات والمواد المنشورة بلغات نادرة وهذه يمكن استخلاصها بتفصيل أكثر من غيرها) ومن جهة أخرى فإن الانتظام في عملية الاستخلاص بتقديم مستخلصات قصيرة جداً يمكن أن يكون إجراء ذاتياً متحيزاً. وعلى ذلك فلا يوجد عادة اثنان من القائمين على الاستخلاص يكتبون المستخلص بصورة متماثلة تماماً، أو حتى اتفاقهم حول المعلومات التى يجب أن يشملها المستخلص.

وأخيراً نجد أنه من المستحيل إعداد مستخلصات إعلامية كاملة لكثير من أوراق البحوث والمراجعات نظراً لأن الوثيقة الأصلية هذه تحتوى على العديد من الأشخاص والأفكار.

(٣) المستخلص الكشفى - الإعلامى:

وهذا النوع من المستخلصات أكثر شيوعاً من كل من الإعلامى أو الكشفى الخالص. وفى حالتنا هذه فإن أجزاء من المستخلصات تكتب بالأسلوب الإعلامى بينما الجوانب الأخرى من الوثيقة والأقل أهمية تعامل بطريقة كشفية، وفى الواقع فإن هذا النوع من المستخلصات يحقق أعلى إمكانية فى بث المعلومات وبأقل طول ممكن.

(٤) المستخلصات النقدية Critical Abstracts

يعتبر هذا المستخلص مراجعة نقدية مركزة Critical Review ومثل هذه المستخلصات النقدية تؤدى نفس الغرض الذى تؤديه مراجعات الكتب النقدية، ولكنها فى حالة المستخلصات تطبق على التقارير ومقالات الدوريات وغيرها من المواد المختصرة، هذا والمستخلص النقدى هو مستخلص تقييمى، أى أن القائم بعملية الاستخلاص يعبر عن وجهات نظره بالنسبة لنوعية عمل المؤلف، وربما يقوم بمقارنته بعمل الآخرين، فقد يشير القائم بالاستخلاص إلى ضعف المنهجية وإلى عدم دقة العينة أو حجمها أو صياغة أسئلة البحث، كما قد يقوم بمقارنة

النتائج بالمسوحات السابقة فى نفس المجال. ومعنى ذلك أنه يجب على القائمين بعمل المستخلصات النقدية أن يكونوا خبراء فى المجال، ومن أمثلة المستخلصات النقدية هذه: مراجعات الرياضيات ومراجعات الميكانيكا التطبيقية.

Mathematical Reviews/ Applied Mechanics Reviews.

ومما سبق نجد أن المستخلصات النقدية تعتبر مستخلصات غير عادية فى الاستخدام، ولكن لها جاذبيتها للمستفيد. فالمستخلص النقدى الجيد لا يصف محتوى الوثيقة فحسب، ولكنه يقيم العمل أيضاً. والمستخلص النقدى يبين عادة مقدار عمق العمل، مع التعليق على كفاية التجربة ومنهجية المسح وكذلك خلفية المستفيدين الذين يفترض أن يوجه إليهم العمل ودلالة الإسهام الفعلى لهذا العمل فى تطور المعرفة.

والمستخلصات النقدية ذات فاعلية خاصة فى إبراز الوثائق ذات الأهمية غير العادية. ومع ذلك فالمستخلصات النقدية نادرة الاستخدام ذلك لأنه لابد أن يتوفر فى القائم بعملية الاستخلاص درجة عالية من التخصص الموضوعى وليس مجرد الفهم العام للموضوع، ذلك لأنه سوف يقيم العمل فى علاقته بالأعمال الأخرى المنشورة فى نفس المجال. وسيدلى القائم بالاستخلاص برأيه وتحليله، وواضح أن مثل هؤلاء الأفراد المتخصصين القادرين على النقد الموضوعى نادرون، ووقتهم ثمين للغاية.

هذا ويذهب كليفيلند (Clevland, D. 1990) فى مقارنته للمستخلص الشارح (أو الواصف) والمستخلص الإعلامى والمستخلص النقدى إلى أن المستخلص الشارح قد يقرر مثلاً أن "عدد براميل النفط التى أنتجت فى السعودية قد جاء فى هذه المقالة" أما المستخلص الإعلامى فيقدم البيانات المحددة، وفى الحالة السابقة فإنه يقرر أن هناك مثلاً عشرة بلايين برميل أنتجت فى عام كذا وهناك أيضاً المستخلص النقدى الذى يقدم لنا حكما قيميا أو تعليقا تحريريًّا على ورقة البحث وفى مثلنا السابق فيقرر المستخلص أن "المقال يوضح لنا عدد البراميل المنتجة ولكنه لا يوضح فى أى عام كان ذلك وبالتالى فالمعلومات تعتبر غير ذات قيمة".

ويلاحظ أن المستخلص النقدى فى معظم الأحيان يكون أقصر من المستخلصات الأخرى ويكتب بمصطلحات عامة، ويعتبره البعض كنوع من أدوات الإحاطة السريعة Alert device ولكنه لا يحل محل ورقة البحث الأصلية على الإطلاق، بل ولا يعتبره البعض كنوع من أنواع المستخلصات (Cleveland, D., 1990, p. 164).

أما المستخلص الإعلامى فيحاول تقديم أكثر ما يمكن من البيانات الكمية والنوعية على قدر المستطاع، وهذا النوع هو أكثر الأنواع فائدة خاصة بالنسبة للبحوث التجريبية. ويقارن البعض المستخلص الإعلامى بالهيكل الإنسانى مثلاً بدون لحم فالناظر إليه يعطى تفصيلات كافية تصلح لإعادة بنائه.

والمستفيد لا يحتاج فى معظم الأحيان مع المستخلص الإعلامى الاطلاع على الوثيقة الأصلية نظراً لأن المعادلات والنتائج الإحصائية وأجزاء من الجداول يشملها المستخلص. ويجب أن يغطى المستخلص الإعلامى ما يلى:

(أ) الهدف ونطاق العمل.

(ب) المناهج البحثية المستخدمة.

(ج) النتائج Results.

(د) الاستنتاجات Conclusions

فالبند الأول هام لأنه يساعد القارئ على تحديد مدى احتياجه لورقة البحث هذه من عدمه دون الاطلاع على بقية المستخلص والبند الثانى يجب أن يغطى التجهيزات المستخدمة وجميع التفاصيل المنهجية حتى يكتسب المستفيد فهماً جيداً عن البحث.

(٥) المستخلص ذو الغرض الخاص Special purpose

(وبعض الباحثين يطلقون على هذا النوع من المستخلصات المستخلصات المتحيزة Bias التى لها غرض معين ولكن ذلك يتناقض مع تعريف المستخلص نفسه الذى لابد أن يكون موضوعياً (Objective) وبالتالى فإن المستخلص ذى الغرض الخاص يصمم للاستجابة لهدف خاص تضعه خدمة الاستخلاص (Borko, H., 1975, pp. 18-20).

أى أن هذا المستخلص يعكس وجهة نظر تهم الجمهور المستهدف، أو أن هذا المستخلص يشمل فقط المعلومات من جزء معين من المقال له أهمية خاصة للجمهور المختار.

وعلى سبيل المثال فإن مستخلصات من مقالات مجلة الجمعية الطبية الأمريكية JAMA والموجودة فى قاعدة معلومات دوريات الصحة هى مستخلصات مكتوبة خصيصاً للشخص العادى وليس الشخص المتخصص. وبالتالى فإن هدف القائم بالاستخلاص هو التركيز على أجزاء من المقالات التى لها تطبيقات عملية مباشرة ويلاحظ هنا أن المستخلصات التى تعد من نفس المقالات تختلف تماماً فى الميدلاين MEDLINE حيث يكون الجمهور المستهدف هنا هم الأطباء والمهنيون.

وإذا كان لنا أن نرجع إلى معايير أنزى (ANSI) فهناك اختبار عام للنوعية والجودة وهى نوع المستخلصات فى قاعدة البيانات فالمستخلصات الإعلامية يجب أن تكون المستخلصات السائدة مع بعض المستخلصات الأخرى (الشارحة والتى تمزج بين الشارحة والإعلامية) فى بعض الحالات.

والمستخلص الموجه لخدمة غرض معين قد يسمى Slanted Abstract كما سبقت الإشارة وهذه يمكن أن تكون مستخلصات إعلامية أو كشفية أو نقدية أو مستخلصات قصيرة، ولكنها جميعاً موجهة نحو اهتمامات جمهور أو مستفيدين معروفين، وبالتالى فهذه المستخلصات تصلح للخدمات المحلية فى شركات معينة أو بنك وذلك عند إصدار نشراتها الاستخلاصية. وعلى سبيل المثال فإذا تم إنشاء وحدة معلومات لخدمة الأفراد الذين يعملون بشركة لتجهيز الأغذية Food processing فإن هذه الوحدة تسعى لتعريف هؤلاء الأفراد بالوثائق التى تتصل بعملهم. وقد تكون عملية تجهيز الأغذية مجرد موضوع جانبى أو ثانوى فى محتوى الوثيقة أو ملحق منها، فإن المستخلص الموجه نحو تجهيز الأغذية والذى يخدم هؤلاء الأفراد يعتبر ذا أهمية بالغة.

(٦) وهناك أيضاً المستخلصات المتحيزة للنتائج Findings oriented Abstracts

وهذه تقترب من النوعين السابقين وتدمجهما ولكن بهدف خاص، فهذه المستخلصات تركز على النتائج التى انتهى إليها البحث أكثر من اهتمامها باختيار المواد طبقاً لاهتمامات جمهور أو مستفيد معين.

وهذه المستخلصات ذات أهمية كبرى للمديرين وغيرهم ممن يحتاجون إلى معلومات سريعة فهى تقدم لهم هذا النوع من المستخلصات بالنتائج أولاً ثم يعود المستخلص إلى استكمال الصورة بعد ذلك مبيناً الأهداف ومناهج البحث والطرق المتبعة.

رابعاً: اتجاهات البحوث والدراسات عن المستخلصات فى التسعينيات :

قامت الباحثة بإجراء بحث إنتاج فكرى على قاعدة بيانات مستخلصات علم المعلومات (ISA) لفترة التسعينيات، وقد قامت بتنظيم المخرجات المتصلة باتجاهات البحوث فى البنود التالية:

(أ) تعليم كيفية الاستخلاص بالهيبرتكست:

تحت عنوان التحدى المنظور كتب الباحث كولتاى (Koltay, T., 1995) عن استخدام الهيبرتكست كأداة تعليمية لطلاب المكتبات والمعلومات، حيث يستخدم حاسب الماكنتوش لتعليم الطلاب مفهوم المستخلص وأنواعه وأغراضه فضلاً عن اكتساب معرفة أكثر عمقاً عن المستخلصات الإعلامية والكشفية ومراحل قراءة وكتابة المستخلصات.

(ب) أنواع جديدة من المستخلصات:

يحتوى الإنتاج الفكرى فى التسعينيات على أنواع مستخلصات لم تذكر من قبل من بينها المستخلصات الوثائقية ووضع نموذج منهجى لإعدادها (Molina, M., 1995) حيث أشار الباحث إلى أن تطوير هذا النوع من المستخلصات يستدعى تضافر مجالات عملية منها اللغويات والمنطق وعلم النفس المعرفى، كما أن نموذج التحول النصى يعتمد على استراتيجيات مجمعة لها مراحل أربعة وهى القراءة والفهم ـــ الاختيار ـــ التفسير ـــ التخليق. كما يحتوى الإنتاج الفكرى على نوع جديد من المستخلصات هى المستخلصات البنائية Structured Abstracts حيث قام هاربورت (Harbourt et al. 1995) وزملاؤه بتحليل أهداف المستخلصات البنائية فى الدوريات الحيوية الطبية فى نظام ميدلاين Medline وذلك للتعرف على إمكانية تطوير الاسترجاع الببليوجرافى.

(ج) الإفادة من التكشيف فى الاستخلاص:

يتناول كوهين (Cohen J., 1995) كيفية استخدام المصطلحات الكشفية الآلية لإعداد المستخلصات، ويسمى الباحث مصطلحات الكشاف التى يتم توليدها الأضواء Highlights ويلاحظ أن بحوث كوهين مازالت فى المرحلة التجريبية ولكنها تتناول لغات عديدة منها الإنجليزية، والأسبانية والألمانية والروسية واليابانية.

(د) أثر الاتصالات عن بعد على الاستخلاص:

يتناول ليرن (learn, L., 1993) تأثير التطورات الحديثة فى الاتصالات عن بعد فى خدمات التكشيف والاستخلاص خاصة بالنسبة لشبكة الإنترنت والشبكة الوطنية للبحوث والتربية وقد وضع الباحث بعض التوصيات والاستراتيجيات فى هذا الشأن.

(هـ) زيادة التطبيقات على الأقراص المكتنزة:

هناك دراسات عديدة تتناول تطبيقات وتطورات استخدام الأقراص المكتنزة، فقد تناول الباحث وودز (Woods, S., 1994) المحتويات الجارية Current contents على الديسكت Diskette حيث تشمل نفس الدوريات وبيانات الكتب كالطبعات المطبوعة ولكن بميزة جديدة وهى مستخلصات المقالات وهذه تضم ديسكات فى كل من العلوم الزراعية والبيولوجية والبيئية وكذلك الطب الإكلنيكى والهندسة والعلوم التطبيقية والعلوم الحيوية والعلوم الطبيعية والكيمياء وعلوم الأرض فضلاً عن العلوم الاجتماعية والسلوكية ويقوم معهد المعلومات العلمية (ISI) بإنتاجها. كما تناول الباحث توميلو (Tomaiualo 1993) مستخلصات الخدمة الاجتماعية على القرص المكتنز حيث يحتوى القرص الواحد على محتويات الملف الببليوجرافى لجميع العاملين فى الجمعية الوطنية للخدمة الاجتماعية، وقد أشار الباحث إلى اعتبارات التجهيزات المادية والبرامج وكيفية البحث والتوثيق.

وعلى نفس الخط فقد تناول توميلو (Tomaiualo N. 1992) كشاف الاستشهادات المرجعية فى العلوم فى طبعته على الأقراص المكتنزة (SCI-CDE)

والتى تحتوى علي قاعدة بيانات متعددة الارتباطات الموضوعية والتحسينات التى تمت عليها مؤخراً من حيث إضافة المستخلصات والكلمات المفتاحية، كما تمت فى هذه الدراسة مقارنة هذا الكشاف (SCI-CDE) بالميدلاين على الأقراص المكتنزة (Medline on CD-ROM).

(و) الاستخلاص بمساعدة الحاسب الآلى:

تناول كرافن (Craven, T., 1993) هذا الموضوع فى مشروع نص تكستنت Textent text network management system حيث وضع بعض الأساليب اللازمة لوزن الجمل ومن بينها: الوزن الإيجابى أو السلبى بالنسبة لجذور الكلمات فى مقاطع مختارة، أى الوزن الذى يتم بالنسبة للقوائم العامة للكلمات المفتاحية Okey Words والوزن الذى يتم بالنسبة لدرجة تكرار الجذور.

(ز) مدى استمرار خدمات التكشيف والاستخلاص التقليدية:

تناولت الباحثة وايلى (Wiley, D., 1994) هذه المشكلة، مشيرة إلى أن العديد من نظم الاسترجاع والبحث على الخط المباشر تواجه خطر التقادم نظراً للتغيرات السريعة فى تكنولوجيات الحاسبات والاتصلات، وقامت الباحثة وايلى بشرح ما قام به بعض المنتمين لقواعد البيانات العلمية الرئيسية لتطوير نظم جديدة للاستجابة للاحتياجات المتغيرة خاصة بالنسبة لاحتياجات الطريق الفائق السرعة Super high way

(ح) المعايير وتقييم المستخلصات فى المجالات المختلفة:

هناك اهتمام أيضاً فى الاتجاهات الحديثة لدراسات المستخلصات، بالمعايير التى يمكن تطبيقها عالمياً على مختلف مجالات العلوم الطبيعية والاجتماعية والإنسانيات، وقد تناول الباحث تيبو (Tibbo, H. 1992) هذا الموضوع حيث قام بدراسة مقارنة لتطبيق معايير (ISO/ ANSI) (وهى معايير المنظمة العالمية للتوحيد القياسى والمعهد الأمريكى للمعايرة) على فئات محتويات من هذه القطاعات العلمية، حيث تبين له أن هذه المعايير تتناسب تماماً مع العلوم الطبيعية ولكنها فى حاجة إلى مراجعة لتتلاءم مع المجالات الأخرى وبالذات الإنسانيات.

خامساً: من الذى يعد المستخلصات:

تكتب المستخلصات عادة بواسطة ثلاث جماعات أساسية هى:

المؤلفون أو خبراء المجالات الموضوعية أو المستخلصون المهنيون، ولعل المؤلف هو المستخلص المثالى لأنه الخبير المعترف به فى المجال لاسيما إذا كان له خبرة سابقة وممارسة فى كتابة المستخلصات لسنوات عديدة، وإذا كان التفكير المبدئى يشير إلى المؤلف كأفضل من يكتب المستخلصات الخاصة به، إلا أن الواقع يشير أيضًا إلى أن المؤلفين ليسوا بالضرورة هم أفضل المستخلصين (وإن كانت المستخلصات التى يكتبها المؤلفون وتأتى ضمن ورقة البحث أرخص فى التكاليف). ذلك لأنهم قد يفشلون فى التركز على الشيء الأهم من الناحية الموضوعية، فقد يلجأ هؤلاء عند كتابتهم للمستخلص وفى ذهنهم الترويج لعملهم وبالتالى سيكون المستخلص غير متوازن بالنسبة للمستفيد وطبقاً لما تذهب إليه معايير المعهد الأمريكى فلا أهمية إذن من الذى يكتب المستخلص، فيمكن أن يكتب بواسطة مؤلف العمل الأصلى أو بواسطة أحد المستخلصين المهنيين، ما دامت هذه المستخلصات قد اتبعت فيها قواعد إرشادية للجودة وكشافات الدوريات العامة التالية تستخدم مهنيين فى الاستخلاص

- Reader's Guide Abstracts.

- Periodical Abstracts on Disc.

- Magazine Article Summaries.

ومن جهة أخرى فهناك العديد من الكشافات التى تكتب بواسطة المؤلفين وتصحب مقالة الدورية فعلى سبيل المثال فإن SCISEARCH بدأت فى إضافة مستخلصات مكتوبة بواسطة المؤلفين منذ يناير ١٩٩١، وإذا لم يتوفر مستخلص المؤلف مصاحباً للمقال المنشور، فإن تسجيلة SCI-EARCH تدخل بدون المستخلص، وهناك حوالى ٥٠% من مدخلات التسجيلات عام ١٩٩١ تشمل المستخلصات (Tenopir, C., 1993, 44).

وقد يستخدم مصطلح Synopsis للدلالة على المستخلص الذى يعده المؤلف، كما أن هناك مصطلحاً بديلاً لهذا المستخلص الخاص بالمؤلف ويمكن تسميته

المستخلصات ذات الموضوع الفردى Homotopic Abstract وهى تنشر مع المقال الأصلى للمؤلف. ومن الناحية النظرية فالمؤلف فى وضع جيد بالنسبة لإعداد مستخلص متجانس Sympathetic Abstract للمقال ومن الممكن تعديل مستخلص المؤلف هذا ليلائم أسلوب خدمة المعلومات، ومن المعروف أن المستخلصات يغطيها حقوق المؤلف، والمؤلف يعارض أحياناً فى نقل المستخلص الذى وضعه للاستخدام فى موضع آخر، ولكن معظم خدمات المعلومات تدرك أن هذا الاعتراض غير جدى لأن ما سيقومون به من الاستعانة بمستخلص المؤلف، إنما هو لصالح المؤلف نفسه.

ومع ذلك فمستخلصات المؤلفين هذه لها مشاكلها النوعية فبعضها مكتوب أو مصاغ بطريقة غير سليمة سواء من ناحية الطول أو القصر فى تسجيل البيانات أو إبراز بعض الجوانب المتميزة التى يراها المؤلف. وعلى ذلك فتستخدم مستخلصات المؤلفين كنماذج ولكن يتم إخضاعها بعد ذلك للتحرير والمراجعة.

(Rowley, J., 1982, p. 9-16)

وقد حذر بوركو وبيرنير (Borko, H., 1975) من الاعتماد على المؤلفين كمستخلصين، ذلك لأن نوعية المستخلصات التى يكتبها المؤلفون ستختلف فيما بينها، نظراً لأنه من غير المتوقع أن يعرف جميع المؤلفين الإجراءات والقواعد اللازمة للاستخلاص، فمطبوع الاستخلاص الذى يعتمد على مستخلصات المؤلفين وحدهم سيعانى من عدم الانتظام عادة فى طول المستخلص وأسلوبه .. وهناك بعض دوريات الاستخلاص التى تعتمد على المتخصصين الموضوعيين Subject Specialists وهؤلاء ليسوا مستخلصين مهنيين ولكنهم مهنيون فى تخصصهم العلمى، وإذا ما تم تدريب هؤلاء ومارسوا كتابة المستخلصات لفترة طويلة فسينتجون مستخلصات ذات نوعية عالية. ولعل مجلة المستخلصات الطبية Excerpta Medica هى نموذج أو مثل رائد فى هذا المجال حيث تعتمد فى إعداد مستخلصاتها على الخبراء فى مجال العلوم الطبية، وإن كان هنا تحفظ منطقى هو أن هؤلاء الخبراء ليسوا متخصصين فى مختلف العلوم الطبية، أى أن عملهم سيكون محصوراً فى تخصصهم الضيق فقط.

أما الجماعة الثالثة فتتكون من المستخلصين المتفرغين، ويمكن القول بصفة عامة بأنه من السهل تدريب المتخصصين الموضوعيين على أعمال الاستخلاص من جعل المستخلصين المهنيين متمكنين من الحقل الموضوعى، ولكن المستخلصين المهنيين ــ كجماعة يمكن أن تنتج مستخلصات عالية الجودة خاصة إذا كان عملهم مركزاً فى أحد الجوانب الموضوعية (Cleveland, 1990) وعادة يتم الاستعانة بالمستخلصين المهنيين بعض الوقت للإفادة من مهاراتهم فى كتابة المستخلصات وأحياناً لخبرتهم الموضوعية فى المجالات العلمية.

ويعتقد بعض الخبراء أن المستخلصين قد يكونون ممتازين فى الاستخلاص وحده دون غيره من بقية أنشطة تحليل الوثائق كالتكشيف وبالتالى فتلجأ بعض الشركات إلى الفصل بين النشاطين، وهذا ما يفعله الناشر ولسن Wilson حيث يحتفظ بقسم الاستخلاص فى كامبردج (ماسا شوستس) لكتابة مستخلصات Reader's Guide Abstracts أما التكشيف فيتم فى المقر الرئيسى فى برونكس Bronx مع توفر الاتصال الإلكترونى بين الموقعين (Tenopir, C. 1992, 15) ولقد تبين ذلك أيضاً لمعهد البترول الأمريكى (API) حيث جاء ذلك فى مقال برينير (Brenner, E, 1989, P.1) الذى جاء فيه:

تعتبر مهارات كل من التكشيف والاستخلاص مختلفة تماماً عن بعضهما، فالمستخلص الجيد ليس بالضرورة مكشفاً جيداً، فالعديد من المكشفين ليس لديهم أدنى مهارة فى الاتصال عن طريق الكتابة. ومع ذلك فهناك من يرى غير ذلك، حيث يمكن أن يقوم نفس الشخص بالاستخلاص والتكشيف فى ذات الوقت، وبالتالى يتم الانتهاء من العمليتين فى وقت قصير فضلاً عن انخفاض التكاليف. أما الباحث لانكستر (Lancaster, F., 1991, p. 105) فيذهب إلى أنها خطوة صغيرة تلك التى تفصل بين مرحلة التحليل المفهومى للتكشيف وبين إعداد المستخلص المقبول، هذا فضلاً عن أن النظام المتبع فى كتابة المستخلص يمكن أن يساعد فى تحديد ما يجب تغطيته فى التكشف وما يمكن حذفه، والحقيقة أن هناك مزجاً فى عملية القراءة والتقاط المفاهيم فى كل من عمليتى الاستخلاص والتكشيف، وبالتالى فهذا ما يبرر دمجهما مع شخص واحد كلما كان ذلك عملياً وممكناً.

وفى الواقع فهذا الدمج فى العمليتين مع شخص واحد، هو الإجراء المتبع فى كل من:

- Magazine article summaries.
- Periodical Abstracts.

سادساً : إجراءات وخطوات القيام بالاستخلاص:

يعرف الاستخلاص بأنه العملية التى يتم بها عمل المستخلصات التى سبق بيان طبيعتها وأنواعها فى الصفحات السابقة، وهناك بعض المبادئ الهامة التى ينبغى اتباعها لإعداد مستخلصات جيدة وهذه المبادئ تزداد أهميتها مع محاولة الاستخلاص الفعلى، ذلك لأن التطبيق العملى يصقل مهارات القائم بعملية الاستخلاص ويحسن من أسلوبه (Rowley, 1982) .

اختيار الوثائق للاستخلاص:

وهذه الخطوة هى بداية برنامج الاستخلاص، ذلك لأنه من العسير بل من المستحيل أن تقوم الهيئة التى تقدم على القيام بالنشاط الاستخلاصى، إعداد مستخلصات لجميع الوثائق التى لها علاقة بالموضوع الذى تهدف الهيئة إلى تغطيته. ولعل المؤشرات السبعة التالية تعتبر أضواء على هذا الاختيار والتمييز بين ما ينبغى استخلاصه وما لا ينبغى استخلاصه من الوثائق:

١- الوثائق التى لها ارتباط وعلاقة باهتمامات المستفيدين.

٢- الوثائق التى تقدم إسهامات جديدة لأحد المجالات.

٣- التقارير النهائية أو غيرها من التقارير التى يدعمها منهجية سليمة وأدلة مقنعة.

٤- الوثائق التى تحتوى على معلومات من الصعب الوصول إليها وذلك مثل الوثائق الأجنبية أو التقارير الداخلية والمذكرات وغيرها من الوثائق ذات التداول المحدود.

٥- الاستعراضات Reviews الهامة.

٦- الوثائق التى تحتوى على معلومات موجودة فى مصادر موثوق بها كالمعلومات المتوفرة فى الدوريات المتخصصة والمهنية المشهورة.

٧- المصادر الصادرة عن مؤسسات معينة وخاصة الدوريات والتقارير، والتى تتكفل الجهة القائمة بالاستخلاص بتغطيتها بشكل شامل.

إن سياسة الاختيار القائمة على الاستخدام تمثل مبدأ سليماً، ولكن تطبيق هذه السياسة يعتمد على التغذية المرتدة Feedback الفعالة من المستفيد إلى القائم بعملية الاستخلاص.

هذا وتبدأ عملية الاستخلاص كما يشير إلى ذلك كليفيلند Cleveland بتقرير صلاحية المادة للاستخلاص، خاصة إذا كان المستخلصون يعملون داخل إطار سياسة معينة بالنسبة لاستبعاد مواد معينة "كالملاحظات والاتصالات القصيرة" notes and short Communications كما قد تقوم بعض هيئات الاستخلاص بفرز المقالات قبل تقديمها للمستخلصين الذين لديهم قواعد مرشدة بالنسبة للتغطية الموضوعية وإن كان للمستخلص فى جميع الأحوال تقييمه وأحكامه ضمن هذه القواعد وهى عادة تدور حول ما يلى:

(أ) الضوابط الاقتصادية:

وهنا يهتم القائمون على الاستخلاص بالأولويات نظراً لأن التكاليف بالنسبة للزمن والإنتاج هامة، وبالتالى فالمواد ذات الأولوية المنخفضة ستستبعد.

(ب) المواد الهامة:

فالملاحظات والاتصالات والخطابات للمحررين لها أهمية متدنية.

(ج) مصدر المطبوع:

وهنا تستبعد المطبوعات التى يصدرها ناشرون ليست لهم شهرة فى مجال النشر فضلاً عن مدى تغطيتها فى خدمة تكشيف واستخلاص مشهورة..

(د) الاهتمامات الموضوعية للمستفيدين:

قد تنجم بعض المشكلات عندما لا تكون الصورة واضحة للمستفيد أمام المستخلص، وعلى ذلك فسياسة القواعد المرشدة تساعد المستخلص على اتخاذ القرارت، ذلك لأن المادة نفسها قد تكون قيمة ولكنها ليست هامة بالنسبة للمستفيد.

خطوات الاستخلاص:

(١) الخطوة الأولى:

وهذه تتمثل فى التسجيل الدقيق للمراجع، ذلك لأن الخطأ فى رصد المرجع سيجعل القارئ يتجه إلى مكان آخر بحثاً عن المعلومات، وعادة يأتى المرجع فى رأس المستخلص وسيساعد ذلك القارئ على الاختيار، فالعنوان شأنه فى ذلك شأن الكشاف يحتوى على كلمات ومصطلحات هامة فى اتخاذ القرار المعنى فى قراءة المستخلص أو عدم قراءته. ولابد من أخذ عناية وتركيز للعناصر التالية:

العنوان: فالعنوان الجيد يمكن أن يكون أداة مفتاحية فى تحديد المعلومات من أجل الاسترجاع، حيث تحمل العناوين الجيدة الكلمات الدالة على المحتوى.

ويمكن للمستخلص أن يضيف للعنوان بعض الكلمات التوضيحية (بين الأقواس) ولكن فى حدود ضيقة حتى لا يؤدى ذلك إلى تضليل المستفيد، وفى حالة أوراق البحوث باللغات الأجنبية، تقوم العديد من هيئات الاستخلاص بترجمتها حيث قد يكتفى بالعنوان المترجم للإنجليزية، وقد تستبقى الهيئة العنوانين.

المؤلف: يرى بعض المستخلصين ضرورة بداية المرجع بالمؤلف لأن ذلك هو المدخل الببليوجرافى السائد ولأن العديد من المستفيدين يعتمدون فى تقييمهم للبحث على شهرة المؤلف فى المجال، كما أن الدخول فى الاستشهادات المرجعية بالمؤلف أكثر يسراً وسهولة، وإن كانت خدمات التكشيف والاستخلاص تفضل البدء بالعنوان على اعتبار أنه يخدم المدخل الموضوعى. وعادة يدخل المؤلف باسم العائلة بوضع باقى المؤلفين حسب الترتيب العادى مع استخدام الحروف الاستهلالية لتوفير المساحة وإن كانت هذه الحروف الاستهلالية قد تكون مضللة فى حالة وجود اسم عائلة واحد لعدد كبير من المؤلفين.

الهيئة التى يعمل بها المؤلف: وهذه تكتب بعد اسم المؤلف مباشرة كعنوان مراسلة للمستفيدين الذين قد يطلبون إيضاحات أو فصلات reprint .

الهيئة الممولة: إذا كانت الوثيقة تقريراً فنياً أو ورقة تعتمد على نتائج البحوث الممولة، فيجب إدراج اسم الهيئة ضمن القسم الخاص بالمرجع وفى هذه الحالة فالمعلومات يجب أن تشمل اسم الهيئة والمنحة ورقم العقد أو الاتفاقية.

مصدر المطبوع: وهذا بالطبع هو الوجدة المفتاحية فى المرجع ويجب أن يكون دقيقاً ومنتظماً واتباع تقانين موحدة وذلك كما يلى: عنوان الدورية (العناوين المختصرة يمكن استخدامها إذا كانت المختصرات معيارية ومفهومة) رقم المجلد/ رقم العدد/ الصفحات الكاملة/ السنة. مثل:

J. Amer. Soc. Inf. Science. 30 (5) 290-95 (1979).

أما الكتب فتشمل العنوان/المؤلف/الناشر/مكان النشر/التاريخ/الصفحات/الثمن.

وهذه هى معظم المطبوعات المشمولة، ولكن قد تكون هناك براءات اختراع فتكتب كما يلى: العنوان/ المخترع/ الهيئة/ البلد الذى صدرت منه البراءة/ رقم براءة الاختراع/ تاريخ نشرها.

وأخيراً فيمكن أن يشمل المدخل المرجعى هذا معلومات عن الوثائق باللغات الأجنبية فيجب بيان اللغة الأصلية، مع مصدر الترجمة إذا كانت قد ترجمت.

(٢) الخطوة الثانية:

وهذه تتمثل فى تحليل المحتوى للوثيقة لأغراض الاستخلاص، وهذه العملية تشبه عملية التكشيف مع وجود اختلاف بينهما، فهدف التكشيف تحديد المفاهيم المفتاحية لإنشاء قائمة كلمات أو واصفات محكومة تدل على محتوى الوثيقة، أما المستخلص فيهدف إلى إنشاء بديل للوثيقة والتى تمثل هيكلها ولكن بكلمات المستخلص فى هذه الحالة.

وهناك خمسة مؤشرات يبحث عنها المستخلص فى إعداده لتحليل المحتوى وهى:

أ – **الأهداف:** أى لماذا كتبت هذه الوثيقة وما هو الهدف الذى كان فى ذهن المؤلف وعادة هذه تكون فى بداية ورقة البحث.

ب – **المنهجية:** الأوراق التى تعكس العمل التجريبى هناك بعض الأساليب الفنية والمناهج المستخدمة والتى وضعها لاسيما تلك المناهج الجديدة، أما بالنسبة للوثائق التى تعكس العمل التجريبى فيجب أن يلاحظ المستخلص مصادر البيانات وكيفية معالجتها خاصة بالأساليب الإحصائية.

ج – **النتائج:** ما هى العلاقات والارتباطات الملاحظة؟ هل البيانات خام، هل تم الحصول على النتائج من قياس واحد أو أن هناك تكراراً للتجربة وماذا عن الصحة والثقة أو دقة النتائج كما يراها المستخلص؟

د – **الخاتمة:** ما هى الفروض المقبولة والمرفوضة؟ ما هى التقييمات والتطبيقات والاقتراحات وعلاقة ذلك بأهداف البحث: هذا والخبير المتمرس بعملية الاستخلاص لا يقرأ كل كلمة فى الوثيقة ولكنه يفرز ويختار الأجزاء الهامة منها. ولعل الفقرات الأخيرة من الوثيقة هى أكثرها تركيزاً على محتوياتها. وكذلك الفقرات المعنونة "النتائج" "التوصيات" "المناقشات" "العمل المستقبلى" تعتبر مصادر ثرية للإفادة منها فى المستخلص.

أما الفقرات التقديمية فهى عادة تهدف إلى تنوير القارئ وربما تكون ذات أهمية بالنسبة لغير المتخصص فى الموضوع.

(٣) الخطوة الثالثة :

وتتمثل هذه فى كتابة المستخلص، وعند هذه النقطة يجب بناء المفاهيم فى سرد قصير، ويجب أن يعبر عن التحليل باللغة الطبيعية، ويلاحظ أن الإطار العام أداة مفيدة عند الكتابة، وهذا الإطار يشمل النقطة الرئيسية ثم نقاط فرعية./ نقطة رئيسية ثم نقاط فرعية .. وهكذا، ويجب ألا تكرر الجملة الأولى فى المستخلص (وهى جملة نقدية) المكتوب فى العنوان، كما يجب أن يكون بناء المستخلص موحداً ومنطقياً أى أن يكون هناك بداية ووسط ونهاية بحيث تترك الجملة الأخيرة فى المستخلص انطباعاً لدى القارئ أنه لم يعد هناك شيء يذكر بعد ذلك، كما يجب أن يتجنب المستخلصون التعبيرات الغامضة أو الطويلة وفى الواقع فإن الجمل

الكاملة ليست ضرورية دائماً، ذلك لأن المقاطع يمكن أن تحمل المعنى، وأخيراً فلا يجب أن يكون المستخلص النقدى شاملاً لوجهة نظر متحيزة.

(٤) الخطوة الرابعة:

يفضل فى هذه الخطوة كتابة اسم القائم بالاستخلاص لإظهار المسئولية، وذلك بالحروف الاستهلالية ويمكن أن يظهر الاسم الكامل فى مكان آخر فى المطبوع المستخلص.

(٥) الخطوة الخامسة:

تتضمن ترتيب المستخلصات ويتم ذلك عادة حسب الترتيب الهجائى للعناوين، كما يجوز الترتيب الهجائى أيضاً حسب المؤلف، أو حسب واصفات الموضوع، وأن يكون الترتيب مصنفاً أو ترتيباً قاموسياً. وفى جميع الحالات تقريباً تحتاج المستخلصات إلى كشافات لدعم الملف الأساسى وذلك هام بالنسبة للمستخلصات المصنفة، وعادة يكون هناك حاجة لكشافات للمستخلصات حسب ما يلى: الموضوعات الأسماء/ عناوين الدوريات، مراجعات الكتب/ التصنيف/ المختصرات. ويتبع فى إعداد الكشافات القواعد الأساسية للتكشيف. وكلما كان التكشيف جيداً، كانت الإفادة من المستخلصات جيدة.

سابعاً : كيفية كتابة وتحرير المستخلص:

يجب تحرير المستخلصات جيداً قبل نشرها، فقد تكون هناك ضرورة لبعض الحذف أو الإضافة أو الجنوح عن السياسة المتبعة، أو أخطاء فى المراجع أو مختصرات لا داعى لها أو تركيب نحوى للجمل غير سليم ... إلخ.

وأول أعمال التحرير هو ضبط القسم المرجعى مع الأوراق الأصلية فعادة يكون هناك بعض التصويبات فى أسماء المؤلفين أو العناوين أو أسماء الدوريات أو المصادر أو المجلد أو العدد، فضلاً عن البيانات الرقمية الأخرى.

أما بالنسبة للنص نفسه فيجب أن يستخدم المحرر مصطلحات معيارية للارتفاع بمستوى المستخلص، والمحرر هو الجسر الذى بين الكاتب والمطبعة، فعملية

التحرير عملية متخصصة ونشاط مهنى ولعل **القواعد المرشدة التالية** تساعد فى تحقيق ذلك (Rowley, 1982) .

(١) يبدأ المستخلص عادة بجملة "قيادية" تلخص أى معلومات أساسية لم يحملها العنوان، وذلك لمساعدة القارئ فى اختيار هذه الوثيقة أو استبعادها.

(٢) يجب أن يقاوم القائم بالاستخلاص إغراء استخدام الجمل الطويلة لتجنب التكرار. فتدفق المعنى وسلامة الأسلوب والقراءة لا ينبغى أن يضحى به من أجل الاختصار. ومن ذلك نجد أن الجملة ذات الاثنى عشرة كلمة هى مستخلص صالح للقراءة.

(٣) يجب أن تكون فقرة المستخلص متماسكة وليس سلسلة من الجمل غير المتصلة ببعضها ويجب أن تكون الجمل كاملة مع عدم حذف الأفعال والمعلومات... إلخ وذلك باستثناء السطر الأول من المستخلص فيمكن ألا يكون جملة صحيحة.

(٤) يمكن أن ترقم النقاط داخل الجملة، ويصلح هذا الترقيم فى المستخلصات الكشفية أكثر من غيرها.

(٥) يجب تجنب الكلمات المضللة أو التى تحمل معانى متعددة فضلاً عن استخدام المختصرات المعروفة جيداً للقارئ مثل سم، كجم وهكذا، وإذا كانت موجهة للمتخصص الكيميائى مثلاً فهناك مختصرات معروفة للكيميائيين وهكذا بالنسبة لغيرهم من المتخصصين.

(٦) الدقة والاختصار من أهم مقومات المستخلص، أى أنه يجب استبعاد الحشو Redundant Phrases كأن يكتب مثلاً: "وقد درس المؤلفون" "فى هذا العمل" ورقة بحث تتضمن النتائج التالية" "وهذا يمكن أن يؤخذ على أنه يبين أن" "وهذه المناقشات تؤدى إلى التوصيات التى" أى يجب عدم استخدام كلمات يمكن إلغاؤها مع استمرار الوضوح مع الاختصار.

(٧) إن الأسلوب وترتيب الأفكار يجب أن تعكس فكر المؤلف، إلا إذا كان هناك ما يبرر تعديل ذلك. هذا من ناحية الأسلوب، أما من ناحية المحتوى فيجب أن يعكس المستخلص الوثيقة الأصلية وما يقصده المؤلف ويركز عليه.

هذا والعديد من الوثائق التى تعد المستخلصات لها، هى بحوث وتقارير ومقالات الدوريات ومستخلصات أوراق البحوث **يجب أن تعكس ما يلى:**

(أ) الهدف من البحث.

(ب) المنهج المستخدم (مع ذكر الأجهزة والأدوات).

(ج) النتائج التى توصل إليها الباحث.

(د) النتائج الفرعية.

ويجب أن يركز المستخلص على المصطلحات الجديدة التى أدخلها المؤلف وعلى النظريات والفروض والنتائج، مع العناية والدقة بالنسبة للبيانات والأرقام.

هذا ويتعلم الشخص منا أن يكون مستخلصاً جيداً من خلال الخبرة العملية فقط، وبالتالى فيمكن هنا أن نسجل بعض القواعد المرشدة فقط، وهى التى تتصل بكيفية قراءة المادة العلمية للتحديد السريع لنقاطها الهامة وقد قام العالم (Cremmins, 1982) بالمناقشة التفصيلية لكيفية قراءة المقال لالتقاط النقاط الأكثر أهمية فيه مع تقديم بعض القواعد التى تخدم هذا الغرض.

ولعل أهم مميزات المستخلص الجيد هى: الاختصار والدقة والوضوح، وبالتالى فيجب على القائم بالاستخلاص تجنب التكرار الذى لا داعى له، وأن يتم بناء المستخلص اعتماداً على المعلومات فى عنوان المادة، وليس تكرار نفس العنوان، كما ينبغى على القائم بالاستخلاص استبعاد المعلومات الأخرى التى غالباً ما تكون معروفة من قبل القارئ، أو التى لا تكون ذات أهمية مباشرة له، وتمثل هذه المعلومات الخلفية التاريخية. ويؤكد الباحثان (Borko & Bernier, 1975) على ضرورة قيام القائم بالاستخلاص ببيان ما قام به المؤلف فعلاً، وليس ما حاول القيام به، وفشله فى تحقيق ذلك، فضلاً عن تجنب ما ينوى المؤلف عمله مستقبلاً.

وكلما كان المستخلص قصيراً كان أفضل ما دام المعنى واضحا ودون تضحية بخاصية الدقة. وهناك بعض الكلمات غير الضرورية والتى يمكن استبعادها كلما أمكن مثل كلمة المؤلف، المقال، فضلاً عن استخدام بعض المختصرات المعيارية مثل ALA جمعية المكتبات الأمريكية.

ويلاحظ أن المستخلصات فى بعض الحقول العلمية والتكنولوجية يمكن أن تستخدم مختصرات عديدة، وعلى الرغم من أن ذلك يوفر مساحة، إلا أنه قد يؤدى إلى زيادة الوقت المطلوب من القارئ حيث قد يضُطر إلى استعارة المقال الأصلى لفهم المستخلص.

وهناك كلمات تحمل معانى متعددة بالنسبة لجماعات مختلفة من القراء وبالتالى فيجب ألا يتضمن المستخلص مثل هذه الكلمات والاقتصار على تلك التى لها معنى واحد بالنسبة لجميع القراء.

ويذهب بعض المستخلصين إلى ضرورة تغيير كلمات المؤلف، إلا أن ذلك قد يؤدى أحياناً إلى تشويه المعنى، من أجل ذلك فقد أكد كوليزون (Collison, 1971) على أهمية استخدام القائم بالاستخلاص المصطلحات المستخدمة بواسطة المؤلف كلما أمكن ذلك. هذا ويوصى المعهد الأمريكى الوطنى للمعايير بعدم استخدام الأفعال كلما أمكن ذلك، وإن كان هذا الاحتياط غير دقيق فى كثير من الأحيان، خاصة وقد أوحى بوركو Borko وغيره باستخدام الفعل الماضى فى وصف الإجراءات التجريبية، واستخدام الفعل المضارع للنتائج المستخلصة من التجارب، ولعل ذلك أن يكون منطقياً، فالأنشطة التى أشار إليها المؤلف هى أشياء من الماضى، أما النتائج فهى التى تعيش معنا فى الوقت الحاضر. هذا وقد وضع الباحث وييل (Weil, 1970) وزملاؤه جدولاً إرشادياً للمستخلصين، وأوصوا بأن يعمل القائم بالاستخلاص من الشيء المحدد إلى الشيء العام (أى أن النتائج لها أولوية، أما التفاصيل فتأتى بعدها، والجمل العامة تأتى فى نهاية المستخلص، وإن كانت هذه التوصية يجب أن تؤخذ بحذر حتى لا تقدم النتائج قبل بيان الغرض من البحث، وفى الواقع فقد أوصى معهد المعايير الأمريكى بأن يبدأ المستخلص بجملة تعكس الموضوع، أى أنها جملة تعكس الهدف الأساسى من الوثيقة.

ولعل أهم مبادئ الاستخلاص تلك الواردة فى الجدولين (١)، (٢) وهى المبادئ التى وضعها مركز التوثيق العسكرى، والتى تتضمن ما ينبغى أن يشمله المستخلص وما لا ينبغى أن يشمله فضلاً عن المصطلحات المستخدمة (Lancaster F.W., 1991, pp. 99-102).

جدول رقم (١)

ما يجب أن يتضمنه المستخلص	ما لا يجب أن يتضمنه المستخلص
* قم بفرز الوثيقة لتحديد الحقائق المفتاحية.	* تغيير معنى أصل الوثيقة.
* وجه المستخلص للجمهور المستهدف.	* التعليق على الوثيقة أو تفسيرها.
* سجل ما تم التوصل إليه فى الدراسة.	* ذكر الأعمال السابقة.
* سجل كيف تم العمل.	* ذكر النتائج التجريبية التفصيلية.
* قم بوضع النتائج مبكراً فى الجملة الدالة على الموضوع.	* وصف تفاصيل الأجهزة العادية.
* ضع التفاصيل فى الجمل التالية.	* ذكر الأعمال المستقبلية.
* قم بوضع البيانات والجمل العامة آخر شيء فى المستخلص.	* استخدام مصطلحات ذات دلالات مختلفة لجماعات مختلفة.
* قم بتمييز الموضوعات عن بعضها.	* ذكر كلمات لا داعى لها لأنها تعكس شيئاً واضحاً.
* قم بالتمييز بين التجربة والفرض.	* قول الشيء نفسه بطريقتين.
* كن إعلامياً فى الفرض ولكن باختصار.	* استخدم المترادفات بشكل كبير.
* تحرَّ الدقة وعدم استخدام الكلمات المختلف عليها.	* استخدم وصفاً موسعاً للأشياء.
* استخدم الجمل القصيرة الكاملة.	* استخدم الأسلوب التلغرافى.
* استخدم الكلمات القصيرة البسيطة المألوفة.	
* تجنب الكلمات غير الضرورية.	
* استخدم المصطلحات الأكثر عمومية كلما أمكن ذلك.	
* استخدم اللغة الإنجليزية الفنية العادية.	
* استخدم الجمل المباشرة.	
* قم بوصف النتائج بالفعل المضارع.	
* استخدم المختصرات حسب الضرورة.	
* قم بالإحالة المرجعية الببليوجرافية الكاملة.	

جدول رقم (٢)

مبادئ الاستخلاص كما نشرها مركز التوثيق العسكرى

الإطار العام

١- قم بإعداد المستخلص الإعلامى كلما أمكن ذلك.

٢- يتضمن المستخلص من ٢٠٠ -٢٥٠ كلمة.

٣- قم باستخدام نفس المصطلحات الفنية المستخدمة فى التقرير.

٤- أن تتضمن المحتويات ما يلى:

هدف الدراسة

طرق القيام بالدراسة

التحقق من صحة النتائج

النتائج

التطبيقات

٥- استخدم الكلمات بدل الجمل كلما أمكن ذلك.

٦- عدم استخدام الرموز أو الحروف النادرة.

٧- عدم استخدام المختصرات غير العادية.

٨- عدم استخدام معادلات أو هوامش.

٩- عدم استخدام بيانات فهرسة وصفية.

ملحق الدراسة رقم (١)

<u>ثامناً : نماذج أنواع المستخلصات</u>

• **العنوان** : قائمة رؤوس موضوعات مكتبة الكونجرس على هيئة مكنز.

المؤلف: Dybstra, Mary

المصدر: Library Journal, Vol. 113, No. 4, P. 42-46, Mar. 1988

مستخلص شارح Indicative (من المصادر التربوية ERIC).

يتناول التغييرات الحديثة فى قائمة رؤوس موضوعات الكونجرس (LCSH).

مبيناً أن تبنى القواعد المعروفة للمكانز قد خلقت ارتباكاً لأن قائمة رؤوس موضوعات الكونجرس ليست مكنزاً حقيقياً، وأوضح التمييز بين رؤوس الموضوعات والمصطلحات ثم أورد مثلاً ممكناً لمشكلات LCSH (٥ مراجع) . (MES)

مستخلص إعلامى (من مستخلصات علم المكتبات والمعلومات) LISA

على الرغم من أن المصطلحات المستخدمة فى قائمة رؤوس موضوعات مكتبة الكونجرس مصطلحات قديمة وغير سليمة، مما أدى بالمستفيدين إلى عدم استخدام الفهارس اليدوية، إلا أنه من الواضح أن المشكلات الأكثر عمقاً تكمن فى الطريقة التى تم بها بناء الرؤوس وكيفية ربطها بعضها ببعض، أما التطورات الحديثة الخاصة بتبنى القواعد الواضحة المستخدمة فى المكانز المعيارية فلا تعتبر خطوة نحو تحسين القائمة، وذلك لأن قائمة الكونجرس ليست مكنزاً، وأى محاولة لجعلها مكنز عن طريق تغيير الأكواد على الرؤوس الحالية وفى التركيب الحالى يعتبر عملاً مضللاً ومستحيلاً فى التحقيق، ومن المهم إدراك التمييز الأساسى بين رأس الموضوع والمصطلح Term لأن هذا التمييز يشكل الأساسى الذى تختلف عنده قائمة رؤوس الموضوعات من المكنز.

المستخلص المزجى Combiation Abstract.

على الرغم من أن المصطلحات المستخدمة فى قائمة رؤوس موضوعات مكتبة الكونجرس مصطلحات قديمة وغير سليمة، مما أدى بالمستفيدين إلى عدم استخدام الفهارس اليدوية، إلا أنه من الواضح أن المشكلات الأكثر عمقاً تكمن فى الطريقة التى تم بها بناء الرؤوس وكيفية ربطها بعضها ببعض، أما التطورات الحديثة الخاصة بتبنى القواعد الواضحة المستخدمة فى المكانز المعيارية فلا تعتبر خطوة نحو تحسين القائمة. وقد تم نقد الطبعة الجديدة لقائمة رؤوس الكونجرس بالتفصيل، ويميز المؤلف بين رؤوس الموضوعات ومصطلحات المكنز ويقدم حلولاً لمشكلات قائمة الكونجرس.

المستخلص النقدى :

يتناول التغييرات الحديثة فى قائمة رؤوس موضوعات الكونجرس (LCSH) مبيناً أن تبنى القواعد المعروفة للمكانز قد خلقت ارتباكاً لأن قائمة رؤوس

موضوعات الكونحرس ليست مكنزاً حقيقياً، وأوضح التمييز بين رؤوس الموضوعات والمصطلحات، ثم أورد حلولاً ممكنة لمشكلات LCSH الجديدة، لقد قامت هذه المقالة بتوليد جدل ونزاع Controversy مع رد من مكتبة الكونجرس بأنهم لم يقصدوا فى الطبعة الجديدة أن تكون مكنزاً حقيقياً، ولكن الجدل ما زال مستمراً، لأن قائمة رؤوس موضوعات الكونجرس تشبه المكنز ولكنها لا تستجيب لمعايير المكانز الحقيقية.

مستخلص الغرض الخاص.

سيتم إثارة المفهرسين بهذا الجدل والحوار المتصل بالسؤال المطروح هل الطبعة الجديدة لقائمة رؤوس موضوعات الكونجرس LCSH هى مكنز حقيقى أم لا، ويذهب المؤلف إلى أن الطريقة التى تم بها بناء القائمة وربط الرؤوس بعضها ببعض يشير إلى مشكلات عميقة بالقائمة، أما التطورات الحديثة الخاصة بتبنى القواعد الواضحة المستخدمة فى المكانز المعيارية فلا تعتبر خطوة نحو تحسين القائمة ذلك لأن قائمة رؤوس موضوعات الكونجرس ليست مكنزاً وأى محاولة لجعلها مكنز عن طريق تغيير الأكواد على الرؤوس الحالية وفى التركيب الحالى يعتبر عملاً مضللاً ومستحيلاً فى التحقيق، ومن المهم إدراك التمييز الأساسى بين رأس الموضوع والمصطلح لأن هذا التمييز يشكل الأساس الذى تختلف عنده قائمة رؤوس الموضوعات من المكنز، والمكشفون والمفهرسون يقومون بنفس العمليات والأنشطة الفكرية، ولكن أدوات كل منهم تختلف بطريقة جذرية وبنائية.

(٢) (ب) ملحق الدراسة رقم (٢)

بعض نماذج من خدمات الاستخلاص فى مختلف المجالات:

اعتمدت الباحثة فى اقتباسها لهذه النماذج على عدة مصادر أهمها (Hurt, 1994)، (أحمد بدر، ١٩٩٢)، (محمد فتحى عبد الهادى، ١٩٩٤) علماً بأن العديد من هذه المستخلصات قد أصبحت ضمن قواعد المعلومات الإلكترونية ويمكن بحثها بالخط المباشر:

- Astronomy and Astrophysics Abstracts.
- Berlin: Springer – verlayg, 1969.

والتغطية هنا دولية مع التركيز على اللغة الإنجليزية، أى أن المستخلصات منشورة بلغات الإنجليزية والفرنسية والألمانية، وهى مصنفة إلى عدد (١٠٨) فئة، وتستخدم المستخلصات التى كتبها المؤلفون كلما أمكن ذلك وإن كانت هناك بعض المداخل بدون مستخلصات ويوجد كشافات بالمؤلفين والموضوعات.

Biological Abstracts, Philadelphia: Biosls 1962.

تغطى حوالى (٩٠٠٠) دورية وسلسلة من أكثر من مائة قطر، وترتب المراجع فى فئات موضوعية عريضة، وهناك أربعة كشافات تسمح بالإتاحة الموضوعية من الاستشهادات المرجعية باستخدام المؤلف/ والتجميع المنهجى الحيوى Bio Systematic وأسماء الأجناس genus فضلاً عن المصطلحات الموضوعية.

Biological Abstracts: Subjects in context (BASIC).

وكذلك كشاف: CROSS

Computer Rearragement of Subject specialties.

Chemical Abstracts. Columbus, Ohio: Amercan chemical Society, 1907.

تعتبر هذه الدورية أكثر دوريات الاستخلاص شمولاً فى العالم وهـى تنشر أكثر من (٥٠٠٠,٠٠٠) مستخلص فى العام ولها كشاف مؤلف وكشاف براءات اختراع رقمى، وكشاف بالكلمات المفتاحية وذلك بالنسبة للإصدرات الأسبوعية، أما المجلدات نصف السنوية فلها كشاف مواد كيميائية وكشاف موضوعات عامة وكشاف معادلات وكشاف نظم الحلقات Ring وكشاف مؤلف وكشاف براءات اختراع.

- Computers And Control Abstacts: London: Institution of Electrical Engineers, INSPEC 1966.

هذه هى الجزء الثالث من مستخلصات العلوم Science Abstracts وقد أصدرت ما يزيد على (٤٠,٠٠٠) مستخلص إعلامى وشارح فى الثمانييات ويوجد كشافات فى كل عدد وذلك بالموضوع والمؤلف والببليوجرافيات والكتب

والمؤتمرات وبراءات الاختراع وهى تصدر شهرية مع تجميعات كل ستة أشهر وكل أربع سنوات.

- Dissertation Abstracts International. Ann Arbor, Mich: University Microfilms Internation al, 1938.s

وهى تصدر شهرياً فى قسمين رئيسيين أولهما (A) فى الإنسانيات والعلوم الاجتماعية وثانيهما (B) فى العلوم والهندسة، ويبلغ عدد رسائل الدكتوراه التى يتم تلخيصها سنوياً حوالى (٣٠,٠٠٠) رسالة وهى تشمل إلى جانب الجامعات الأمريكية جامعات فى كندا وبريطانيا وأوروبا. والقائمة الرئيسية مرتبة هجائياً حسب المجال الموضوعى ثم بواسطة الجامعة، وهناك أيضاً كشاف مؤلفين وكشاف موضوعات، وهذه تعتبر نموذجاً جيداً للمستخلصات التى يعدها المؤلفون، وقد أصبحت هذه المستخلصات منذ عام ١٩٨٠ متاحة على الخط المباشر شهرياً من خلال نظام لوكهيد وهيئة BRS .

- EXCERPTA MEDICA: Amsterdam, Netherlands, 1947-

هذه خدمات استخلاص رئيسية للإنتاج الفكرى الطبى، واهتمامها الأساسى بالطب الإنسانى والعلوم الأساسية المساعدة، وإن كان هناك تغطية لبعض العلوم الطبية القريبة كالتمريض، وهى تنشر فى أقسام ويغطى كل قسم تخصصاً معيناً ويوجد لها كشافات مؤلفين وموضوعات ويتم ترقيمها سنوياً.

FERTLIZER ABSTRACTS, Muscle Fertilizer Shoals, Ala: Tennessee valley Authority, National Fertilizer. Development center, Technical library, 1968.

هذه الدورية هى مقال جيد لدورية مستخلصات فى مجال متخصص وتنتشر أساساً لخدمة جماعة بحثية تخدم رسالة معينة Mission والمستخلصات تظهر فى ثلاثة أجزاء رئيسية هى: التكنولوجيا والتسويق والاستخدام.

Historical Abstracts, Santa Barbara, Calif. Clio press, 1955.

يتم إعداد المستخلصات بواسطة الباحثين والمتخصصين فى المجال ويتم توقيع المستخلصات بالقائمين بها، وتتبع الدورية ترتيباً مصنفاً مع كشافات سنوية للمؤلفين والتراجم والجانب الجغرافى والموضوعى.

MATHEMATICAL REVIEWS. Province R.I: American Mathematical Society, 1970.s

هذه مثل جيد للمستخلصات النقدية وموقعة بواسطة المراجعين. والتعلقيات النقدية فى التوضيحات نموذجية، والتغطية تشمل كلاً من الرياضيات البحتة والتطبيقية، وهى مرتبة بالموضوع مع جدول تفصيلى للمحتويات حيث تظهر فيها الموضوعات المدرجة، وهناك كشافات للمؤلفين وكذلك للكلمات المفتاحية.

PHYSICS ABSTRACTS. London: Instiution of Electrical Engineers, 1898.

تعتبر هذه الأداة نموذجاً للتغطية العريضة إذ تشمل كل فروع الفيزياء، كما تشمل الببليوجرافيات والكتب وبراءات الاختراع وأعمال المؤتمرات والتقارير والدوريات والرسائل العلمية، والترتيب الأساسى حسب التصنيف الموضوعى مع وجود كشافات مؤلفين وموضوعات وكشافات لمختلف أنواع المواد المغطاة وتتميز المستخلصات بأنها مختصرة وواصفة وليست تقييمية وتعتمد على المؤلفين إلى حد كبير.

PSYCHOLOGICAL ABSTRACTS. Washington, D.c.
American Psychological Association, 1927.

وتشمل التغطية هنا مقالات الدوريات والكتب والتقارير الفنية والمستخلصات موقعة لكل مادة مشمولة، والترتيب مصنف حسب المجالات الرئيسية لعلم النفس ثم يقسم فرعياً إلى موضوعات أضيق، وكل عدد يوجد به كشاف مؤلفين وكشاف موضوعات.

مصادر المعلومات المستخدمة فى اختيار النماذج السابقة :

١- أحمد بدر. (١٩٩٢) مصادر المعلومات فى العلوم والتكنولوجيا – الرياض: دار المريخ، ٣٩١ص.

٢- محمد فتحى عبد الهادى (١٩٩٥) المصادر المرجعية للمعلومات فى العلوم الاجتماعية – القاهرة: دار غريب للطباعة والنشر والتوزيع، ٣٩١ص.



الفصل السابع

التحليل المقارن لمصطلحات ومستخلصات العلوم الطبيعية والاجتماعية والإنسانيات على ضوء معايير الاستخلاص وفى بيئة استرجاع إلكترونية وتقليدية

تقديم

تربط نظم الاسترجاع الببليوجرافى المستفيدين بالوثائق عن طريق مضاهاة بدائل الوثائق Surrogates بأسئلة الباحثين، وبدائل الوثائق هذه تشمل عادة العناوين والمصطلحات المحكومة (مثل قوائم رؤوس الموضوعات أو المكانز) ومصطلحات اللغة الحرة والمستخلصات. وتمثل المصطلحات المستخدمة فى جميع هذه البدائل أهمية محورية بالنسبة لكفاءة الإتاحة الموضوعية Subject Access.

وتتناول الدراسة التى بين أيدينا فى البداية مقارنة لهذه البدائل مع بعضها بالنسبة للاسترجاع الجيد، ومقدرة هذه البدائل على تمثيل الوثائق المصدرية بدقة مع تمييز هذه الوثائق عن غيرها فى نفس المجال، وقد أظهرت الدراسات المختلفة فى الإنتاج الفكرى تفوق المستلخصات الإعلامية informative abstracts كأهم عناصر نقاط الإتاحة فى استرجاع المعلومات على الرغم من تكاليف إعدادها العالية.

ثم تتناول الدراسة بعد ذلك بالتحليل المقارن مدى دقة المصطلحات فى كل الإنسانيات والعلوم الاجتماعية والطبيعية بالنسبة للإتاحة الموضوعية لهذه المجالات فى نظم المعلومات، وعلى الرغم من المقولة السائدة فى الإنتاج الفكرى من صلابة مصطلحات العلوم الطبيعية Hard terms وليونة مصطلحات العلوم الاجتماعية والإنسانية Soft terms أو أن العلوم الطبيعية تستخدم لغة اصطناعية تسمح بإنتاج تعاريف دقيقة وفئات محددة، وأن الإنسانيات تستخدم اللغة الطبيعية حيث يكون لكل مصطلح فيها مدى واسع من المعانى .. إلا أن الدراسة التى بين أيدينا قد أظهرت جوانب تحليلية مضيئة بالنسبة لكل من الإنسانيات والعلوم الاجتماعية تساعد على إنشاء نظم استرجاع أكثر فاعلية، فضلاً عن زوال الحواجز بين العلوم الطبيعية والاجتماعية بالنسبة للاستشهادات المرجعية.

وأخيرا تورد الدراسة تحليلا مقارنا لمستخلصات العلوم الطبيعية والاجتماعية والإنسانيات وذلك على ضوء معايير الاستخلاص الدولية والأمريكية (/ISO ANSI) حيث تبين أنه على الرغم من أن هذه المعايير تعكس ممارسات الاستخلاص الجارية فى العلوم الطبيعية، إلا أن هذه المعايير لا تتفق مع نموذج الاستخلاص فى حقل التاريخ (كممثل للإنسانيات)، وتنتهى الدراسة ببعض النتائج والتوصيات المتصلة بالدراسات والبحوث المستقبلية خصوصا بالنسبة للإنتاج الفكرى باللغة العربية.

أولاً : المستخلصات أهم عناصر الإتاحة فى استرجاع المعلومات:

قام الباحث بيرن (Byrne, J., 1975) بدراسة للتعرف على المزايا النسبية للبحث بواسطة العناوين ورؤوس الموضوعات والمستخلصات ومصطلحات اللغة الحرة، وتوليفة من كل هذه العناصر، وقد استخدمت قاعدة بيانات كومبندكس للإنتاج الفكرى الهندسى (Compendex) فى هذه الدراسة، نظرا لاحتوائها على جميع عناصر البيانات ذات الأهمية واللازمة لاسترجاع التسجيلات المطلوبة ..

وقد أظهرت نتائج التجارب التى قام بها بيرن Byrne ما كان متوقعا من أن العناوين وحدها تعتبر غير كافية للاسترجاع الجيد، لأن العناوين وحدها لا تحمل معلومات كافية تسمح للباحث بالقيام بالتحديد الإيجابى للمراجع الصالحة، وقد أشار الباحث إلى أنه يمكن عمل حل وسط باستخدام العناوين ورؤوس الموضوعات مع بعض المصطلحات المساعدة من اللغة الحرة، أما فى حالة البحث بواسطة المستخلصات وحدها أو إدخال كلمات المستخلصات فى الملف المقلوب، فإن تكاليف البحث تزيد بدرجة ملحوظة للحد الذى لا تحتمله معظم نظم الاسترجاع، كما أظهرت نتائج الدراسة أن توليفة العناوين والمستخلصات اقتربت من ١٠٠٪ استرجاع، كما وصل البحث بواسطة المستخلصات وحدها إلى نفس النسبة تقريبا، **أى أن المستخلصات تعتبر أهم عناصر نقاط الإتاحة فى استرجاع المعلومات على الرغم من تكاليف إعدادها العالية.**

ولعل المستخلص الإعلامى من النص الحر يعتبر أقوى الوسائل التى يمكن أن تلتقط البناء الجدلى للوثيقة والمصطلحات الفتاحية فضلا عن تقديم صورة كلية

لمحتوياتها .. وعلى ضوء هذه الأهمية للمستخلص فقد حدد كل من المعهد الأمريكى للمعايرة وهيئة التقييس الدولية (ANSI / ISO) ما ينبغى أن يشمله المستخلص الجيد في المعيار المشترك الذى وضعه كل من ANSI, ISO حيث تزود هذه المعايير بتعليمات خاصة بالنتائج والتى تشكل عادة أهم جزء فى الوثائق .. إذ تذهب هذه التعليمات إلى ما يلى فى إعداد المستخلص:

قم بعرض النتائج باختصار وبطريقة إعلامية كلما أمكن ذلك، فقد تكون هذه النتائج تجريبية أو نظرية، وقد تكون بيانات مجمعة أو علاقات أو ارتباطات correlations أو تأثيرات تمت ملاحظتها .... إلخ، قم بتوضيح القيم الرقمية وهل هذه ترصد لأول مرة بناء على الدراسة أو أنها مستقاة من مصادر أخرى، وهل هذه الأرقام هى نتيجة ملاحظة واحدة أو أنها نتيجة قياسات متكررة. وإذا كانت النتائج كثيرة للغاية فلابد من عمل أولويات تبعاً لما يلى: الأحداث الجديدة والمحققة، النتائج التى تتعارض مع النظريات السابقة، أو النتائج التى يعرف المؤلف أنها متصلة بمشكلة عملية .. وأخيرا فحدود الدقة والثقة والصحة يجب أن تبين

(ANSI, 1979b, p.9; ISO. 1982, p. 25).

لقد استخدم الباحثون واختصاصيوا المعلومات المستخلصات منذ زمن بعيد، للوصول إلى أحكام متصلة بمجالهم وذلك عندما كان استرجاع المعلومات محدودا بالمصادر المطبوعة .. كما ساعدت هذه البدائل (العناوين/ المستخلصات/ المصطلحات الكشفية ...) فى نشر بيانات كافية بطريقة مختصرة ودقيقة، فضلا عن تحقيق متطلبات الإحاطة الجارية (Clevelend, 1990).

ومع دخول نظم الاسترجاع المحسب والتى تتيح البحث الحر فى النص أصبحت المستخلصات تقدم للباحثين إمكانيات جديدة، ذلك لأن هذه المستخلصات تقدم مصطلحات بحثية ونقاط إتاحة أكثر بالنسبة لكل تسجيلة فى قاعدة البيانات (Fidel, 1986) وهذه الإمكانيات الجديدة ذات أهمية بالغة لمجالات الإنسانيات حيث لا تعكس العناوين محتويات الوثائق فى كثير من الحالات، كما أن التكشيف فى قواعد البيانات هذه قد يكون ضحلا (Garfield, 1980) وبالإضافة للعناوين، فإن المستخلصات التى يعدها المؤلفون يمكن أن تعكس بصورة أفضل المصطلحات الجارية فى المجال، وهذا الإجراء أفضل كذلك من مجرد اختيار

مصطلحات أكثر من لغة تكشيف مسبقة. وأخيرا فالمستخلصات تعتبر أدوات قوية للمستفيدين سواء فى وقت البحث أو فى التطبيقات المستقبلية.

ولكن فيضان الإنتاج الفكرى المعاصر مع حاجة الباحث الحديث لأدوات الإحاطة الجارية قد ضخم من الصعوبات التى تكتنف إعداد المستخلصات، ومن بين تلك الصعوبات اضطرار خدمات التكشيف والاستخلاص إلى العمل بمعاونة المتطوعين وغيرهم حيث يكتب المؤلفون أو المتطوعون معظم المستخلصات، على أن يقوم المكشفون بوضع الواصفات الموضوعية لتسجيلات الوثائق اعتمادا على هذه المستخلصات ..

وحتى مع استخدام المكانز لتوجيه أنشطة المكشفين، فهناك قليل من ضبط الجودة بالنسبة للمستخلصات، وإذا ما أغفل المستخلص بعض المفاهيم، فإن المكشف لن يستطيع استعادتها، ولن تتمثل هذه المفاهيم فى أى مكان آخر فى بدائل الوثيقة.

ويلاحظ أن المشكلات السابقة تختلف عن مشكلات المصطلحات والتى يمكن السيطرة عليها باستراتيجيات البحث وأساليبه .. أى أن الحذف أو الخطأ بالنسبة لتمثيل المحتوى يعتبر أمراً غير مرئى وفى أحيان كثيرة يستحيل استعادته.

ومما سبق يتضح لنا زيادة الأهمية المعاصرة للكتابة الجيدة للمستخلصات فضلا عن ضرورة التحليل الأكثر اتساعا لبناء ومحتوى المستخلص.

ومع ذلك فتفرض الوظائف العديدة للمستخلصات قيدان على البدائل Surrogates (العناوين/ المستخلصات/ المصطلحات الكشفية..)، وأول هذه القيود أن هذه المستخلصات يجب أن تكون مختصرة للدرجة التى يستطيع معها المستفيدون التعرف على العديد منها فى وقت قصير .. ولعل ذلك يتناقض مع ضرورة أن تكون المستخلصات "إعلامية" على قدر المستطاع .. والملاحظ أن المعيارين (ASNI Z39-14-1979 & ISO 214-1976 E) للمعهد الأمريكى والمنظمة الدولية، يؤكدان على أهمية المحتوى المعلوماتى فى المستخلص وعلاقته بالوثيقة الأصلية وذلك كما يلى:

أن يكون المستخلص إعلاميا كما يسمح به نوع وأسلوب الوثيقة، أى أنها يجب أن تمثل علـــى قـــدر المستطـــاع المعلومات الكمية والكيفية التى تحتويها الوثيقة (ANSI, 1979b, p7; ISO, 1982, p.23) أما كليفيلند (Cleveland, 1990) فيستخدم كلمات مخلفة قليلا ولكنها بنفس المعنى "يلخص المستخلص المحتويات الأساسية لتسجيله معرفية معينة، وتعتبر بذلك بديلا حقيقيا للوثيقـــة" أمـــا راولـــى (Rowley, 1982, p.9) فيضيف فكرة الأسلوب حيث يشير إلى أن المستخلص "تمثيل مختصر ودقيق لمحتويات الوثيقة بأسلوب مشابه للوثيقة الأصلية".

وواضح أن عملية الاستخلاص تتحدد عن طريق التجاذب والشد بين ضرورة الاختصار والحاجة للحفاظ على أكثر المعلومات من الأصل على قدر الإمكان والنتيجة هى ما يسمى بفن أو علم الاستخلاص حيث لابد من أن تختار العناصر الأساسية من الأصل على صعوبة تحقيق ذلك من الناحية العملية.

وإلى جانب المحتوى المعلوماتى هذا، هناك جانب آخر يتمثل فيما تقوله الوثيقة Aboutness .. فقد يبنى المؤلفون أوراق بحوثهم حول البناء التقليدى (المقدمات/ المنهجية/ النتائج/ أقسام المناقشة) ولكن اختلافات كبيرة يمكن أن تحدث بالنسبة لهذا البناء .. ذلك لأن هذا الشكل وأن كان يعتبر الشكل المتعارف عليه فى البحوث العلمية وأخيرا فى البحوث الاجتماعية، إلا أنه ليس الشكل الذى يستخدمه بطريقة منتظمة المشتغلون فى الإنسانيات، فهم نادرا ما يضعون أقساما للنتائج أو حتى للمنهجية. وبدلا من ذلك فإن محتويات الكتابات الإنسانية تظهر خالية من الأقسام المعيارية هذه .. وهذا ما يعكسه فحص جميع المقالات فى مجلة (AHR) Americen Historical Review منذ عام ١٩٨٣ وحتى عام ١٩٨٥ فليس هناك أى مقالة تحتوى على (المنهجية أو النتائج) .. ولعل ذلك هو ما جعل هناك شك فى إمكانية كفاية المعايير العالمية لاستخلاص جميع أنواع الإنتاج الفكرى (Tibbo. H.R., 1992).

ثانياً: استخدام المصطلحات فى استرجاع المعلومات والاستشهادات المرجعية فى كل من العلوم الطبيعية والاجتماعية والإنسانيات:

٢-١ الإتاحة الموضوعية فى الإنسانيات ومدى دقة مصطلحاتها:

٢-١-١ وجهات النظرية المتعارضة:

تعتبر صعوبة الإتاحة الموضوعية فى الإنسانيات وعدم دقة لغة المشتغلين بالدراسات الإنسانية، موضع تعليق العديد من الباحثين، فقد ذهب الباحث لانجردج (Langridge) أنه بينما تستخدم العلوم الطبيعية لغة اصطناعية تسمح بإنتاج تعاريف دقيقة وفئات محددة، فإن الإنسانيات تستخدم اللغة الطبيعية حيث يكون لكل مصطلح فيها مدى واسع من المعانى.

أما الباحث فوسكت (Foskett, D.J., 1971) فقد قارن بين عدم دقة مصطلحات الإنسانيات وعدم دقة مصطلحات العلوم الاجتماعية حيث أشار إلى أن علماء العلوم الاجتماعية لا يعبرون عن أفكارهم بمصطلحات مقبولة للجميع وإن كان هناك علماء فى هذه العلوم يستخدمون مصطلحات دقيقة ولكن عدم الدقة ملحوظ أكثر فى الإنسانيات والفنون، حيث يزعم كل كاتب أن من حقه وصف عمله بمصطلحاته الخاصة.

ولكن وجهة النظر هذه لفوسكت قد عارضها الباحث موريس لاين (Line, M., 1974) حيث ذهب إلى أن المواد الأساسية الموضوعية للإنسانيات تتضمن أحداث وأشخاص ومنتجات فنية، وبالتالى فإن المحور الأساسى للمجال يتكون من عناصر يسهل تكشيفها واسترجاعها، أما العلوم الاجتماعية فهى تهتم بالناس الذين يتفاعلون مع بعضهم ومع التغييرات فى الجماعات والمجتمعات، وبالتالى فإن تصميم نظام معلومات يصف هذه الظواهر يتضمن بالضرورة مصطلحات غير ثابتة أو مستقرة.

وأمام وجهات النظر المتناقضة هذه فقد قام ويبرلى (Wiberley,S., 1983) بدراسة لزيادة فهمنا عن الإتاحة الموضوعية فى الإنسانيات حيث تعتبر المصطلحات أساسية للمشكلة كلها، أى أن ويبرلى درس مصطلحات الإنسانيات عن طريق تحليل المصطلحات المستخدمة كمداخل فى الموسوعات والقواميس

المشهورة فى مجال الإنسانيات، وقد تم تجميع هذه الكتب المرجعية وتصميمها لتغطى أكثر الموضوعات أهمية فى مجالاتها، وبالتالى فهى تزودنا بتغطية شاملة لمصطلحات الإنسانيات.

ونظرا لأن المصطلحات التى تم تحليلها فى هذه الدراسة أتت من أكثر من ستين قاموساً وموسوعة مختلفة وغطت مجالات مختلفة أيضا داخل الإنسانيات، فإن هذه المصطلحات تشمل كل جزء من الكلام أو الحديث وتراوحت فى الطول فى كلمة واحدة إلى عدة كلمات .. ولكن معظمها كان مكونا من كلمتين أو ثلاث والأغلبية الساحقة منها كانت "للأسماء" أو "للأسماء المركبة" nominal compounds.

٢-١-٢ فئات مصطلحات الإنسانيات:

يمكن الإشارة لمستويات دقة فئات هذه المصطلحات فى الإنسانيات فى الجدول التالى:

المستوى الأول (الأكثر دقة)	أسماء الأعلام المفردة	
المستوى الثانى	مصطلحات اسمية متعددة	
المستوى الثالث	مصطلحات اسمية عامة (الفئات المتفرعة تشمل جوانب جماعية، ثقافية جغرافية، ايديولوجية وهيئات)	مصطلحات دارجة (الفئات المتفرعة تشمل جوانب مجردة، حسية، وحركية kinetic

وأكثر التقسيمات الأساسية فى النحو التقليدى هو بين أسماء الأعلام proper names والأسماء الدارجة common names، حيث يعرف بعض الكتاب اسم العلم بأنه يحيل إلى واحد فقط من الأشياء وبالتالى فهم يعتبرون أسماء الأعلام أكثر دقة من الأسماء الدارجة (Mario, p. 1954) ومع ذلك فلأغراض استرجاع المعلومات هناك مستويات مختلفة من الدقة بين أسماء الإعلام، فضلاً عن أنه من المستحيل

التمييز بين دقة بعض أنواع المصطلحات الخاصة بالأعلام وتلك الخاصة بالأسماء الدارجة، وعلى كل حال فإن أكثر مصطلحات الإعلام دقة هى أسماء الإعلام المفردة أى أن الاسم هنا يعكس كيانا متفرداً لشخص أو عمل إبداعى واحد وهذه يمكن أن تحدد بالنسبة للزمان أو المكان ..

أما المستوى الثانى من الدقة فهى المصطلحات الاسمية التى يمكن عدّها enumerable proper names وهذه تدل على جماعات يمكن عدّ أعضائها بصفة كاملة كقائمة من أسماء الأعلام المفردة .. وبالتالى فإن المصطلحات الاسمية هذه تعتبر خطوة معزولة من مصطلحات الأعلام المفردة، ولكنها موصولة ومرتبطة بها بطريقة مباشرة.

أما المستوى الثالث فيتضمن المصطلحات الاسمية العامة والدارجة general proper and common terms .. والمصطلحات التى تقع فى الفئتين الأخيرتين تعبران عن معانٍ عديدة range Of meanings وتستخدم بالنسبة لعدد من الكيانات غير المحددة. وغالبا ما يكون من المستحيل التمييز بين الأسماء العامة والدارجة .. ونظرا لأن الفئتين يعتبران على نفس مستوى الدقة، فقد تم تجميعها فى الجزء الكمى لدراسة ويبرلى Wiberley .. وقد اتخذ ويبرلى عناية خاصة فى وضع فئات لمصطلحاته وذلك بالنسبة لما يلى: (أ) التمييز بين أسماء الناس الحقيقيين والتى تمثل التعبير عن مصطلحات الأعلام المفردة، وأسماء الأشخاص فى القصص أو الأسماء الخرافية أو الأسطورية. (ب) تحديد أى المصطلحات الجماعية التى تعتبر مصطلحات اسمية يمكن عدّها. (ج) أن يقوم بوضع أى مصطلح يمكن أن يصنف فى أكثر من فئة واحدة، فى فئة الأقل دقة.

٢-١-٣ تحليل عينة مصطلحات الإنسانيات:

قام ويبلى بتحليل عينة المصطلحات التى اختارها من موسوعات وقواميس الإنسانيات وكانت العينة ممثلة عن طريق تكوين قائمة أساسية من كل من:

- Rogers, A.R. **The Humanities: A Selective Guide to Information Sources** (1979)
- Sheehy, E. **Guide to Reference Books,** 1976 (1980 Suppl.)

ونظرا لأن مصطلحات الأعلام المفردة والمصطلحات الاسمية العامة والدارجة تعتبر محورية فى هذه الدراسة فقد استبعد أى مرجع تنقصه هذه المصطلحات من العينة.

وقد جاءت الأعداد والنسب المئوية لفئات مصطلحات المداخل فى الكتب المرجعية للإنسانيات فى الجدول التالى:

	حجم العينة	أسماء الأعلام المفردة	مصطلحات اسمية متعددة	مصطلحات اسمية عامة ودارجة	مصطلحات أخرى
الأعداد	٦٧٨	٣٩٣	٧	٢٧٣	٥
النسب المئوية للعينة		٥٨,٠	١,٠	٤٠,٣	٠,٧

ويتضح من الجدول أن أسماء الأعلام المفردة هى أكثر المصطلحات فهى تكون حوالى ثلاثة أخماس المجموع .. وهذه النسبة العالية ترفض مقولة أن الإتاحة الموضوعية فى الإنسانيات عسيرة جداً بالضرورة .. ومع استخدام مثل هذه المصطلحات كلما أمكن فإن المكشفين يمكنهم إنشاء نظم استرجاع فعالة .. هذا والمصطلحات الاسمية التى يمكن عدّها والدارجة تشكل حوالى $\frac{2}{5}$ من العينة وهذه المصطلحات غير دقيقة ذلك لأن تعريفها يتميز بالتغيير عبر الزمن وبالتالى فلها مدى واسع من المعانى ولعل أولئك الذين قاموا بالشكوى من صعوبة الإتاحة الموضوعية فى الإنسانيات كانت هذه المصطلحات الأخيرة هى التى تشغل ذهنهم.

٢-٢ مقارنة ممارسة الاستشهادات المرجعية فى كل من العلوم والإنسانيات:

يذهب الباحث المشهور يوجين جارفيلد (Garfield, E. 1980) إلى أن استرجاع المعلومات فى الفنون والإنسانيات يختلف عنه فى العلوم وذلك لأن الوثائق المطلوبة تختلف من وجوه عديدة، ومع ذلك فقد نجح معهد المعلومات

العلمية فى تطويع نظام تكشيف الاستشهادات الأساسى الذى سبق استخدامه فى كشاف استشهادات العلوم (SCI) للاستجابة للاحتياجات المعلوماتية للباحثين فى مجال الإنسانيات وإصدار كشاف الاستشهادات فى الفنون والإنسانيات (AHCI) .. وهناك تأثيرات متوقعة للكشاف الجديد (وقاعدة بياناته الناتجة)، وتتمثل هذه التأثيرات فى تنمية البحوث المتعددة الارتباطات interdisciplinary فضلا عن إتاحة البيانات الببليوجرافية لبحوث الإنسانيات للدراسات الاجتماعية والتاريخية بالإضافة إلى توليد مناهج موضوعية لتقييم دوريات الإنسانيات.

هذا ويهتم استرجاع المعلومات فى العلوم بالأعمال الحديثة نسبيا، فقد أظهرت دراسة جارفيلد (Garfield, E. 1980) أنه فى مجالات الفيزياء والكيمياء على سبيل المثال هناك أكثر من ٦٠٪ من الاستشهادات هى لأعمال منشورة خلال السنوات الخمس الأخيرة. ولكن فى مجال التاريخ هناك فقط ٣٨٪ من الاستشهادات هى لمقالات منشورة خلال السنوات الخمس الأخيرة.

كما أشار جارفيلد أيضا إلى أنه قام بنشر قائمة لثلاثمائة مؤلف فى مجال العلوم كانوا أكثر المؤلفين الذين تم الاستشهاد بهم بين عامى ١٩٦١ - ١٩٧٦ .. وكان أقدم هؤلاء على القائمة عالم ولد عام ١٨٩٩م، وبالمقابل ففى دراسة مشابهة فى مجال الإنسانيات واعتمادا على البيانات الخاصة بمائة مؤلف تم الاستشهاد بأعمالهم فى دوريات الفنون والإنسانيات فى الفترة من ١٩٧٧ - ١٩٧٨ .. فقد تبين أن هومر Homer وهو أقدم المؤلفين قد عاش حول القرن التاسع قبل الميلاد، وهناك خمسة عشر مؤلفاً عاشوا قبل عام ١٤٠٠ بعد الميلاد، وهناك ستة وثمانون مؤلفا ولدوا قبل عام ١٩٠٠م.

وهناك بعض الخصائص الأخرى التى تميز استرجاع المعلومات فى الإنسانيات وهى بروز الاهتمام بالكتب أكثر من الدوريات كمصادر للمعلومات، وكذلك فإن هذه المصادر لا تتقادم ..

وفى مقارنته بين العلوم والعلوم الاجتماعية والإنسانية فى استرجاع المعلومات بواسطة العنوان انتهى جارفيلد (Garfield, 1980) فى دراسته إلى أنه باستخدام كشاف استشهادات العلوم (SCI) فلم يستدعى الاسترجاع أى إضافات للعنوان، ذلك

لأن عناوين معظم المقالات العلمية كانت معبرة عن المحتوى، أما فى العلوم الاجتماعية فقد كانت هناك نسبة أعلى من العناوين غير المعبرة ولكنها ليست كافية أى أنها لا تدعو إلى إضافات للعناوين، ومع ذلك فقد احتوت الفنون والإنسانيات على عدد كبير من العناوين غير المعبرة، مما استدعى قيام بعض المتخصصين بقراءة المقالات وإضافة أسماء أشخاص وأماكن ومفاهيم ونظريات .. إلخ وهذه هى موضوعات تلك المقالات .. ومما يذكر أنه فى عام ١٩٧٨ قام هؤلاء الأخصائيون بإثراء حوالى (٥٠٠٠) عنوان .. والمعروف أن إثراء العناوين أى الإضافة إليها هى عملية تستغرق وقتا طويلا كما أنها عملية مكلفة ولكنها تزيد من فائدة وفاعلية كشاف الاستشهادات الخاص بالفنون والإنسانيات (AHCI).

٢-٣ مقارنة الاسترجاع فى كل من العلوم الطبيعية والاجتماعية :

يذهب الباحث ليفى (Leavy, M., 1983) فى دراسته بعنوان زوال الحواجز Obliteration of Barriers بين العلوم الطبيعية والاجتماعية اعتمادا على بيانات الاستشهادات المرجعية، إلى أنه بمقارنة درجة حداثة مراجع الدوريات الجارية فى كل من العلوم الطبيعية والاجتماعية. فقد تبين له أن الوسيط median لمنتصف الحياة half-life يصل إلى حوالى ست سنوات فى الإنتاج الفكرى لكل منهما .. وهذه النتيجة تتعارض مع وجهة النظر الكهنية Kuhnian الخاصة بالعلوم الاجتماعية (kuhn, T.S, 1970, p. 245) والتى تشير إلى أن الأعمال القديمة يتم الاستشهاد بها فى العلوم الاجتماعية أكثر مما هو الحال مع العلوم الطبيعية .. وقد وضع ليفى عدة تفسيرات لهذه النتائج منها أن العمل الأقدم سيمحى من الاستشهاد نظرا لأن علماء العلوم الاجتماعية يرونه مشمولا فى الأعمال التالية أو لأنهم يرفضون هذا العمل القديم فى فترات التحول الثورى فى الصيغ التى تعبر عن الظواهر الاجتماعية .. وهناك تفسير آخر وهو أن علماء العلوم الطبيعية والاجتماعية قد نسوا أو أنهم لم يقرءوا العمل القديم نهائيا. وفيما يلى تفصيل لبعض الجوانب الأساسية فى هذا المجال.

٢-٣-١ النظرة التقليدية لنماذج استخدام المراجع فى العلوم الاجتماعية:

الاعتقاد الراسخ لدى علماء المعلومات والمؤرخين للعلم وعلماء العلوم الطبيعية والاجتماعية على السواء، هى أن مصطلحات العلوم الاجتماعية أكثر ليونة Softer من العلوم الطبيعية وأقل استمرارية فى تطورها، فالكيمائيون يرجعون أساسا للأعمال التى تعكس "الأفق البحثى" ولا يتم الاستشهاد بالأعمال القديمة لا لأنها عديمة الفائدة بل لأن الباحثين يفترضون أنها معروفة .. وهناك من الأدلة البحثية التى تثبت هذه المقولات خصوصا قبل السبعينيات فقد قارن الباحث برايس (Price, D. 1970) بين نماذج الدوريات المستخدمة فى كل من العلوم الطبيعية والاجتماعية، وقد أظهرت نتائج دراسات برايس أن دوريات الإنسانيات تقع أقل من الوسيط Median، وأن دوريات العلوم الطبيعية أعلى من الوسيط وأن دوريات العلوم الاجتماعية تحوم حول الوسيط hoovered around وقد اعتبر برايس أن الحداثة وعدد المراجع مؤشرات لصلابة hardness العلم وأن هذه النتائج التى انتهى إليها تؤيد وجهة النظر هذه وإن كان برايس نفسه قد وجد أن (كشاف برايس PI) الذى يقيس به مختلف قطاعات العلوم يختلف كثيرا داخل قطاع العلوم نفسه، إذ تراوح هذا الكشاف بالنسبة لعلوم الحاسب من (٣١-٧٧) وكذلك فإن عناوين العلم البيولوجى أظهرت وجود مراجع قديمة بالمقارنة بعناوين العلم الطبيعى .. وبالتالى فقد ذهب كل من جلبرت وولجار (Gilbert, N, 1974) إلى أن "مزاعم برايس بأنه اكتشف مؤشرا لقياس "صلابة Hardness العلم لا يجب اعتبارها أمراً مقررا ولا مناقشة فيه".

٢-٣-٢ العمر الجارى للمراجع فى كل من دوريات العلوم الطبيعية والاجتماعية:

يعتبر العمر الوسيط Median age لمقالات الدوريات المستشهد بها من الدوريات المكشفة فى كشاف استشهاد العلوم (SCI) وكشاف استشهادات العلوم الاجتماعيـــة (SSCI) محورا أساسيا فى اختبار النظرة التقليدية السابقة ..

ويمكن فى البداية إجراء المقارنات الخاصة بكشاف برايس (PI) للمراجع منذ عام ١٩٧٨ وحتى عام ١٩٨١ بالنسبة لآلاف الدوريات المصدرية فى هذين الكشافين .. حيث يلاحظ انخفاض تدريجى فى كشاف برايس (PI) بين الدوريات

المكشفة في كشاف استشهادات العلوم الاجتماعية وذلك خلال الفترة موضع الدراسة (١٩٧٠ - ١٩٧٧) (Miwa, M. 1980, p.148) والجدول التالى يبين لنا استمرار ذلك الانخفاض حتى عام ١٩٨١ وذلك بالنسبة لإجمالى المراجع فى مجموعات دوريات العلوم الطبيعية والعلوم الاجتماعية على السواء (مع فروق ضعيفة).

السنوات	كشاف استشهادات العلوم الاجتماعية	كشاف استشهادات العلوم
١٩٧٨	٤١,١	٤٦,٢
١٩٧٩	٤٠,٧	٤٥,٩
١٩٨٠	٤٠,٥	٤٥,٦
١٩٨١	٣٩,٩	٤٤,٨

كشاف برايس لكل من دوريات كشاف استشهادات العلوم الاجتماعية والعلوم الطبيعية (النسبة المئوية ٪ لإجمالى الاستشهادات المنشورة فى السنوات الخمس السابقة)

ومع صدور دورية تقارير الاستشهادات (JCR) Journal of citation Reports فى منتصف السبعينيات للدوريات المكشفة فى كل من كشاف استشهاد العلوم وكشاف الاستشهاد فى العلوم الاجتماعية، أصبح من الممكن إجراء مقارنات على نطاق واسع لمنتصف الحياة للدوريات المستشهد بها، ذلك لأن منتصف الحياة الوسيط للعلوم الطبيعية بصفة تجميعية وكذلك للعلوم الاجتماعية تجميعيا للدوريات المكشفة اصبح ذا أهمية بالغة، ويظهر ذلك فى الجدول التالى للسنوات ١٩٧٨ - ١٩٨١ مع النسب المئوية للعناوين فى كل قائمة تكون فيها "منتصف الحياة" أكبر من عشر سنوات (Leavy, M, 1983, p.29.).

	١٩٧٨	١٩٧٩	١٩٨٠	١٩٨١
* الدوريات المستشهد بها فى كشاف استشهاد العلوم ودورية تقارير الاستشهادات (حزمة منتصف الحياة)	٢٣٣٠	٢٦٢٨	٢٦٦٨	٢٧٨٢
والعدد وسيط منتصف الحياة (بالسنين)	٥,٨	٥,٩	٦,٠	٦,١
الدوريات التى تتميز بمنتصف حياة أعلى من عشر سنوات (نسبة مئوية)	١٣	١٥	١٥	١٦

الدوريات المستشهد بها فى كشاف استشهاد العلوم الاجتماعية ودورية تقارير الاستشهادات (حزمة منتصف الحياة)	٥٧٨	٥٨١	٥٧٤	٥٣٨
والعدد وسيط منتصف الحياة (بالسنين)	٦,١	٦,١	٦,٢	٦,٤
الدوريات التى تتميز بمنتصف حياة أعلى من عشر سنوات (نسبة مئوية)	١٧	١٤	١٥	١٦

"منتصف الحياة" الراجعة لكل من دوريات العلوم الطبيعية والعلوم الاجتماعية المستشهد بها بين عامى ١٩٧٨ - ١٩٨١

وخلاصة هذا كله أن الدليل الجارى لا يؤيد وجهة النظر التى تقول بأن علماء العلوم الطبيعية يستشهدون بالأعمال الأكثر حداثة من علماء العلوم الاجتماعية (على الأقل إذا كانت مراجع الدوريات فى الدوريات هى المحور).

ثالثا: التحليل المقارن لمستخلصات العلوم الطبيعية والاجتماعية والإنسانيات:

٣-١ المنهجية

قامت هيلين تيبو (Tibbo, H, 1991, p.38) باختيار أربع عينات كل منها (٣٠) وحدة اثنتان منها فى حقلى الكيمياء التحليلة وعلم النفس التطويرى واثنتان من التاريخ الأمريكى (احداهما تحتوى على بدائل المقالات والأخرى من مستخلصات الرسالات Dissertation Abstracts) وقد تم اختيار عينات المستخلصات الخاصة بالكيمياء وبعلم النفس بطريقة عشوائية من بين خمس دوريات فى هذه الحقول يضع معهد المعلومات العلمية (ISI) أعلى رتبة تأثير highest impact عام ١٩٨٥ .. وذلك للتأكد من أن الاختيار قد تم لمستخلصات من مقالات محورية فى هذه المجالات.

أما بالنسبة لمستخلصات التاريخ فقد تطلبت مسلكا آخر ذلك لأن معظم دوريات التاريخ لا تتضمن مستخلصات للمقالات ... وبالتالى فقد تم اختيار المستخلصات من قاعدة بيانات ABC- Clo's America: History and Life (AHC) .. أما العينة الخاصة بالرسالات التاريخية فقد أخذت من قاعدة الأقراص المكتنزة CD-ROM لمستخلصات الرسالات الدولية Dissertation Abstracts

International (Vol. 44-46) مستخدمة كود الواصفات (التاريخ الأمريكى)، وتم اختيار عينة عشوائية من (٣٠) وحدة ..

وباستثناء مستخلصات ABC-Clio فإن جميع المجموعات احتوت على البدائل التى قام بتوليدها المؤلف، وقد استخدمت هذه المستخلصات الأولية (أى التى أعدها مؤلفو المقالات) وليس المستخلصات الموجودة فى الخدمات الثانوية (مثل Psychological Abstract) وذلك حتى تتمكن الباحثة من فحص ومقارنة إسهامات الاستخلاص التى تمت بواسطة المؤلفين .. وواضح صعوبة مقارنة مستخلصات مؤلفين فى كل من العلوم والتاريخ.

والخطوة التالية للاختيار هى معالجة الكلمات word processed مجزأة حسب الجمل المكونة لها ومطبوعة فى نموذج جملة – جملة وأمامه الكود المقابل، وقد قام المؤلف وخبير فى الموضوع لكل مجال بتكويد المحتوى و(أو وظائف) كل جملة مستخدما صيغة معدلة من فئات "محتويات الوثيقة" التى وضعها (ANSI / ISO) (انظر الملحق). وهذه الفئات تشمل (أ) خلفية أساسية (تشير معايير ANSI إلى الخلفية ولكنها تذهب إلى ضرورة وجودها لفهم المستخلص)؛ (ب) النطاق و(أو) الغرض، (ج) الفروض التى يتم اختبارها (د) منهجية البحث (هـ) النتائج (و) عدم وجود أى عنصر من السابق (NOTA).

ويلاحظ أن الفروض لم تكن فئة من فئات ANSI الأصلية، ولكن تم شمولها على اعتبار أن عددا من المستخلصات سيشمل بيانا متصلا بالفروض الخاصة بالبحث وإن كان ذلك لم يحدث فى أى حالة من حالات مجموعات المستخلصات وبالتالى فإن وجود هذه الفئة كان له تأثير قليل على نتائج الدراسة. وأدخلت الملاحظة (NOTA) حتى تكون الفئات شاملة.

وقد كانت الثقة فى التكوين عالية نسبيا لجميع العينات (الكيمياء ٦٠٪ ، علم النفس ٧١٪ التاريخ ٧٤٪، رسالات التاريخ ٧٣٪).

٣-٢ نتائج دراسة الباحثة تيبو Tibbo

٣-٢-١ التركيب اللغوى للمستخلصات:

لقد تبين أن مجموعات مستخلصات مقالات الكيمياء وعلم النفس متشابهتان تماما بالنسبة لعدد ونسبة الكلمات والجمل التى تحتويها تلك المستخلصات، وفى المجموع الكلى فإن المستخلصات الكيميائية احتوت على أكثر من ٤٠٪ من الجمل فى ثلاث مجموعات المستخلصات المقالات، و٣٩٪ من الكلمات .. أما بالنسبة لمستخلصات علم النفس فاحتوت على ٣٦٪ من الجمل، ٣٩٪ من الكلمات .. وفى مقارنة واضحة فإن مستخلصات مقالات التاريخ تمثل فقط ٢٤٪ من الجمل، ٢٢٪ من الكلمات فى عينات مستخلصات المقالات الثلاثة (انظر جدول I).

وكانت الفروق الواضحة بين مستخلصات الرسالات والمجموعات الأخرى، ولم يكن ذلك مفاجئا نظرا لأنها بدائل للكتب ومستخلصات الرسالات وفيها (٣٨٩) جملة، تمثل أكثر من ٤٤٪ من الجمل فى جميع المجموعات، و٤٦٪ من العدد الكلى للكلمات لجميع المستخلصات (انظر جدول Π).

وإذا أخذنا إجمالى هذه المستخلصات فقد بلغت أكثر من ضعف أى مجموعة أخرى، وعلى الرغم من اختلاف طول المستخلص فإن عدد كلمات العنوان التى تحمل محتوى كان متشابها لجميع المجموعات ولكن الكيمياء كانت تعكس عدد (٣٩١) أى (٢٩٪) وعلم النفس منخفض إلى (٣٠١) أى (٢٢٪).

أما إحصاءات مستوى المستخلص فقد أظهرت خصائص متشابهة، فكان المتوسط والرقم المتوسط للكلمات لكل مستخلص متشابه تماما داخل كل مجموعة وذلك بين عينات الكيمياء وعلم النفس، ولكن المستخلصات داخل كل مجموعة اختلفت بطريقة واضحة بالنسبة للعدد الكلى للكلمات التى تحتويها (انظر جدولΠ).

وكان متوسط الأرقام والأرقام الوسيطة Median وكلمات الجملة فى كل عينة متشابهة سواء فى الداخل أو عبر المجموعات، كما أن عدد كلمات كل جملة كان أعلى لكل مجموعة، وكان ذلك واضحا تماما بالنسبة لمستخلصات مقالات التاريخ حيث كانت هناك وحدتان من هذه العينة تحتوى فقط على جملة واحدة فى كل منها (انظر جدولIII) أما الجدول IV فيبين لنا الرقم التقديرى للكلمات التى

تحتويها الوثائق الأصلية (المقالات والرسالات) وهو يكشف اختلافات واضحة فى طول كل مجموعة، وهذا يبين الحاجة الملحة للاختلافات فى البدائل Surrogates من ناحية طولها أو تفاصيلها .. كما يربط الجدول طول الوثائق المصدرية بمستخلصاتها بواسطة حساب "نسبة الضغط" Compression ratio لكل عينة .. وهذا الرقم يساوى الرقم المتوسط لكلمات كل مستخلص فى كل عينة مقسوما على الرقم التقديرى للكلمات فى كل وثيقة فى عينة وثائق المصدر المقابل. وهذه الأرقام تعكس قدرا كبيرا من الاختلافات بين المجموعات، حيث تمثل المستخلصات الكيميائية حوالى ٥٪ من طول الوثيقة الأصلية، ومستخلصات علم النفس ٣٪، ومقالات التاريخ ٠,٩٪ ورسالات التاريخ ٠,٤٪ فقط، وعلى الرغم من أن هذه العينات مازالت صغيرة حتى يمكن التعميم منها، إلا أن الأرقام الناتجة تشير إلى فرصة أكبر للتمثيل الجيد لمقالات الكيمياء أكثر من أى مجموعة من مجموعات التاريخ.

(الجدول I)

الإجمالى والنسب المئوية للجمل والكلمات فى عينات المقالات

	العدد	عينات المقالات ٪	جميع العينات ٪
الجمــل فــى كــل عينــة			
الكيمياء	١٩٧	٤٠	٢٣
علم النفس	١٧٦	٣٦	٢٠
مقالات التاريخ	١١٥	٢٤	١٣
رسالات التاريخ	٣٨٩	–	٤٤
الكلمات فــى كــل عينــة			
الكيمياء	٤,١٣٧	٣٩	٢١
علم النفس	٤,٢١٧	٣٩	٢١
مقالات التاريخ	٢,٣٨٥	٢٢	١٢
رسالات التاريخ	٩,٢٤٦		٤٦
كلمات العنوان فى كل عينة			
الكيمياء	٣٩١		٢٩
علم النفس	٣٠١		٢٢
مقالات التاريخ	٣٤٧		٢٥
رسالات التاريخ	٣٣٥		٢٤

(الجدول II)

كلمات المستخلص لكل عينة

	الكيمياء	علم النفس	مقالات التاريخ	رسالات التاريخ
العدد الكلي للكلمات	٤١٣٧	٤,٢١٧	٢,٣٨٥	٩,٢٤٦
العدد المتوسط للكلمات	١٣٧	١٤١	٨٠	٣٠٨
العدد الوسيط للكلمات	١٣١	١٣٣	٧٧	٣٢٧
العدد المودال Modal للكلمات	–	١٣١	٦٦	٣٣٠
العدد الأكبر للكلمات	٢٩١	٢١٠	١٣٠	٤١٠
الكلمات الأقل للكلمات	٦٦	٧٣	٤١	١٩٢
مدى أرقام الكلمات	٢٢٥	١٣٧	٨٩	٢١٨

(الجدول III)

كلمات الجملة فى كل مستخلص لكل عينة

	الكيمياء	علم النفس	مقالات التاريخ	رسالات التاريخ
العدد الكلي للكلمات	٤,٢٣٧	٤,٢١٧	٢,٣٨٥	٩,٢٤٦
العدد المتوسط للكلمات	٢١	٢٥	٢٥	٢٤
العدد الوسيط للكلمات	٢١	٢٤	٢٢	٢٥
العدد الأكبر للكلمات	٣٢	٤٦	٩٢	٤٠
الكلمات الأقل للكلمات	١٤	١٧	١٢	١٨
مدى أرقام الكلمات	١٨	٣٠	٨٠	٢٢

(الجدول IV)

الرقم التقديرى للكلمات التى تحتويها الوثائق الأصلية

	الكيمياء	علم النفس	مقالات التاريخ	رسالات التاريخ
العدد المتوسط للكلمات	٢,٩٦٣	٤,٣٧١	٨,٥٤٥	٨٨,٨٠٩
العدد الوسيط للكلمات	٢,٩٣٠	٤,٠٠٠	٨,٣٢٢	٨٣,٨٣٤
أكبر الأعداد للكلمات	٥,٨٤٣	٩,٦١٤	١٤,٨١٢	١٤٧,٤٨٨
أقل الأعداد للكلمات	١,٠٧٣	٢,٧٧٣	٤,٤٣٥	٣٩,١٥٦
مدى أرقام الكلمات	٤,٧٧٠	٦,٨٤١	١٠,٣٧٧	١٠٨,٣٣٢
الرقم المضغوط Compression	٠,٠٤٧	٠,٠٣٢	٠,٠٠٩	٠,٠٠٤

٣-٢-٢ المحتوى الوظيفى للمستخلصات ومعايير ANSI/ISO :

تؤيد البيانات المستقاة من تحليل وظائف الجمل التى تحتويها مجموعات المستخلصات الأربعة الفروض التى تشير إلى أن معايير ANSI/ISO ترتبط أكثر بمحتوى الإنتاج الفكرى العلمى وحقيقة المستخلصات العلمية أكثر مما هو الحال فى الكتابات الإنسانية كما يأتينا الدليل من التاريخ.

هذا ويمثل الجدول IV تحليلا لأكواد المحتوى والنسب المئوية لكل فئة داخل المجموعات .. وهذه البيانات تعكس الأكواد النهائية التى تم الوصول إليها بعد أن قام المكودون ببعض التعديل للاختلافات الواضحة. وقد تبين أن إجمالى التكويد أكبر من العدد الإجمالى للجمل بالنسبة للعينة وذلك عند تحليل الأكواد المخصصة لأى جمل خارج إجمالى هذا الكود أو مضافة إليه .. .

وعند النظر إلى جدول V فسنجد جميع الأكواد النهائية – فيما عدا تلك المتعلقة بعينة الكيمياء وهى ٩٪ – تمثل فئات ANSI/ISO المقترحة والباقى يشكل مواد الخلفيات Back ground .. وصورة عينة علم النفس مماثلة للكيمياء .. ويلاحظ أن الخلفيات/الفروض/ تتميز عن المعلومات غير الواردة فى ANSI/ISO حيث تصل إلى عدد (١٩) أى (١٠٪) من الأكواد المخصصة. وفى هذه الحالة لم ير المكودون أى جمل تقع ضمن فئة NOTA (غير الوظائف المحددة كالخلفيات/ والهدف والفروض ..).

ولقد تغيرت سمات الأكواد الخاصة بعينات مقالات ورسالات التاريخ بدرجة كبيرة، وكان الفرق الأكبر فى فئة NOTA، ويوضح لنا جدول V أن ٦٤٪ من جميع الأكواد المخصصة لمستخلصات مقالات التاريخ كانوا بين NOTA .. ولكن هذا التخصيص قل بنسبة ١٪ بالنسبة لعينة رسالات التاريخ حيث إن عدد (٢٨٥) NOTA يساوى (٦٣٪) من الأكواد المخصصة لهذه المجموعة.

ولكن الأهم من هذا، هو أنه على الرغم من أن مستخلصات مقالات التاريخ تصل إلى (٢٤٪) فقط من الجمل فى مجموعات مستخلصات المقالات الثلاث، إلا أن هذه البدائل Surrogates كانت تحتوى على (٩١,٧٪) من جميع الجمل فى هذه المجموعات وتوصف بأنها لا تتوافق مع أى الفئات الخاصة بـــ ANSI/ISO .

ومما يلاحظ أن هناك (٧) من بين (٣٠) مستخلص كيميائى يحتوى على جمل خلفيات Background ، وهى نسبة أعلى من كل من مجموعات مستخلصات مقالات التاريخ وعلم النفس (الجدول VI)، وقد يظن البعض أن مستخلصات التاريخ ستكون هى التى تحتوى على مواد الخلفيات، ولكن ذلك لم يثبت فى هذه الدراسة .. ولعل تفسير ذلك يعود إلى أن مستخلصات مقالات التاريخ قصيرة جداً وبالتالى فلم تتسع لمثل هذه الخلفيات. كما يلاحظ أن هناك (١٧) من مستخلصات رسالات التاريخ الطويلة، تحتوى على بيانات خلفيات .. أما بالنسبة للكيمياء فجميع مستخلصات الكيمياء تحتوى على بيانات عن الهدف ونطاق الدراسة، وهذه علامة تعكس التنظيم الجيد للمستخلصات حيث تُعلم القارئ بطبيعة وحدود وهدف البحث، ولكن الغريب فى هذه المستخلصات أن العديد من البينات المذكورة يأتى فى الجمل الأخيرة من المستخلصات وليس بين الجمل الأولى، أما بالنسبة لعلم النفس فهناك فقط عدد (٢٠) من مستخلصات علم النفس التى تحتوى على مثل هذه المواد، بينما يوجد فقط عدد (١٢) مستخلص مقالات للتاريخ يحتوى هذه المعلومات العامة.

ويبين لنا هذا التحليل بوضوح أن مفهوم المنهجية (والإعلام عنها) يعتبر أكثر أهمية فى المجتمعات العلمية أكثر منه بالنسبة للمشتغلين بالتاريخ.

ومن بين (٢٨) مستخلص كيميائى، (٢٧) بدائل علم النفس يتوفر بيانات عن المنهج، أما بالنسبة لمستخلصات مقالات التاريخ فهناك فقط عدد (٥) وبالنسبة لمستخلصات رسالات التاريخ هناك (١١) فقط تدلنا على المنهج، وبالتالى فإن مستخلصات رسالات التاريخ تحتوى على نسبة مئوية أكبر تحتوى على ذكر للمنهجية مما هو الحال فى مستخلصات مقالات التاريخ.

والفرق بين مستخلصات العلوم والتاريخ كان واضحا فى نشر بيانات النتائج حيث كانت هناك (٢٣) من مستخلصات الكيمياء وعدد (٢١) من مستخلصات علم النفس التى تحتوى على مثل هذه البيانات، كما أن مستخلصات العلوم تعكس فئات المحتوى التى وضعها كل من (ANSI/ISO) بينما لا تحتوى مستخلصات التاريخ إلاّ قليلا من هذه الأنواع.

(الجدول V)

عدد الجمل ونسبتها المئوية لكل فئة فى كل عينة

فئــات المحتــــوى	الكيمياء		علم النفــس		مقالات التاريخ		رسالات التاريخ	
	ن	٪	ن	٪	ن	٪	ن	٪
الخلفيات	٢٢	٩	٥	٣	٧	٥	٢٦	٦
الهدف/ النطاق	٦١	٢٤	٢٠	١٠	١٢	٩	٣٨	٨
الفروض	صفر	صفر	٦	٣	٢	١	٧	٢
المنهجية	٧١	٢٨	٥١	٢٥	٧	٥	٢٧	٦
النتائج	٥٦	٢٢	٨٥	٤٣	٣	٢	٣	١
ما انتهى إليه conclusion	٤١	١٦	٢٦	١٣	١٩	١٤	٦٧	١٥
غير كل الذى سبق NOTA	صفر	صفر	٨	٤	٨٨	٦٤	٢٨٥	٦٣
المجموع	٢٥١	٩٩	٢٠١	١٠١	١٣٨	١٠٠	٤٥٣	١٠١
الجمل	١٩٧		١٧٦		١١٥		٣٨٩	

(جدول VI)

عدد المستخلصات التى تحتوى كل منها كل كود الفئات لكل عينة

	الكيمياء	علم النفس	مقالات التاريخ	رسالات التاريخ
الخلفيات	٧	٤	٦	١٧
الهدف / النطاق	٣٠	٢٠	١٢	٢٢
الفروض	صفر	٥	٢	٤
المنهجية	٢٨	٢٧	٥	١١
النتائج	٢٣	٣٠	٣	٣
ما انتهى إليه conclusiona	٢١	٢٢	١٤	٢٤
غير الذى سبق NOTA	صفر	٧	٢٦	٣٠

٣-٢-٣ معايير ANSI/ISO والمستخلصات التاريخية:

تدلنا البيانات السابقة على أن هناك اختلافات كبيرة فى محتوى مستخلصات التاريخ عند مقارنتها بالبدائل الممثلة للكيمياء وعلم النفس، وعلى الأخص فإن عدد الجمل التى تقع فى فئة NOTA (الوظائف غير المحددة ضمن الخلفيات /

الفروض ..) من عينات التاريخ تشير إلى أن معايير المستخلصات التى وضعتها كل من (ANSI/ISO) قد لا تكون هى مرشدات البدائل المثالية لهذا المجال.

كما أن فهم طبيعة الجمل الخاصة بالمستخلصات التاريخية والتى لا تتلاءم مع أى واحدة من فئات (ANSI/ISO) هو أمر حاسم لفهم محدودية هذه المرشدات وذلك بالنسبة للإنتاج الفكرى غير العلمى وبالتالى فى كيفية تطوير أى مشروعات استخلاص مستقبلية.

هذا ويلاحظ أن العديد من جمل التاريخ التى جاءت ضمن فئة (NOTA) كانت جُملاً وصفية، وبعضها الآخر كانت روائية narrative، كما أن هناك قطاعا كبيرا منها يمثل آراء المؤلفين أو المستخلصين .. وبعض الجمل فى العينتين اللتين أخذتا للتاريخ (من المقالات والرسالات) كانت ذاتية أى معبرة عن وجهة نظر المؤلف .. وبعض هذه الجمل فى النهاية كانت شبيهة جدا بالنتائج التى يصل إليها العلماء. وإن كانت النتائج وما انتهى إليه الباحث conclusions لا تعكس بوضوح طبيعة البيانات الموجودة فى نهاية المقالات والأمر هنا يختلط مع النتائج الاستنباطية. Deductive findings.

وهناك فئات أخرى من المواد التى ظهرت فى المستخلصات التاريخية، ومن بينها ما يقوم به المشتغلون بالتاريخ كثيرا من وضع عملهم فى سياق تاريخى historiographical يأخذون منه ما يدعم البحوث السابقة أو ما يهاجمون به هذه البحوث.

وكما هو الحال فى المستخلصات لعلم النفس فهناك جمل فى البدائل التاريخية تناقش الوثائق المصدرية وليس مناقشة البحوث أو المحتوى التاريخى، وأخيرا فهناك نوع آخر من البيانات لا تتصل مباشرة بالمنهجية ولكنها تضع قوائم بمصادر البيانات أو وصفها.

٣-٣ مناقشات دراسة تيبو Tibbo

٣-٣-١ هل هناك معيار عالمى للاستخلاص؟

أثمرت الدراسة التى بين أيدينا عن عدة **نتائج أولها:** أظهر تحليل المحتوى أن هناك اختلافاً واضحاً بين المستخلصات العلمية والتاريخية، سواء تلك التى

يعدها المؤلفون أو المستخلصون. وقد احتوت مستخلصات العلوم على نسبة عالية من تحديد الهدف والنطاق والمنهجية وبيانات النتائج طبقا لتعليمات كل من (ANSI/ISO) بينما تحتوى المستخلصات التاريخية سواء المقالات أو الرسالات على استنتاجات وخاتمة conclusions (أحيانا فروض أو رأى المؤلف) وكذلك بيانات لا تتلاءم مع كل فئات (ANSI/ISO).

ثانيهما: تدلنا هذه البيانات أن المؤلفين أو المستخلصين فى مجال التاريخ لا يتبعون معايير الاستخلاص الوطنية أو الدولية .. وفى نفس الوقت فإن المؤلفين الذين يتعاملون مع الإنتاج الفكرى العلمى ــ على الأقل فى مجال الكيمياء وعلم النفس ــ ينتجون بدائل تتلاءم مع تعليمات وإرشادات كل من (ANSI/ISO).

وتشير هذه الملاحظات إلى الاستنتاج الذى يؤكد على أن معايير (ANSI/ISO) تتلاءم مع الكتابات العلمية بالنسبة لإعداد المستخلصات. ولكنها لا تتلاءم مع حقيقة الإنتاج الفكرى فى مجال التاريخ أو بصفة عامة فى الكتابات غير العلمية..

ونظرا لتجانس النتائج من هذه المصادر فإن هذا التأكيد يذهب إليه الباحث بثقة على الرغم من صغر العينة النسبى. فعلى سبيل المثال فلم تحتو المستخلصات الكيميائية على جمل لا تأخذ فى اعتبارها معايير (ANSI/ISO)، بينما يقع حوالى ثلثى جمل المستخلصات التاريخية خارج تعليمات(ANSI/ISO)، هذا فضلا عن أنه رغم وجود بعض الفروق، إلا أن كلا من مستخلصات الرسالات التى أعدها المؤلفون والمستخلصات المكتوبة بواسطة المستخلصين لقاعدة بيانات ABC- CI كانت متشابهة تماما بالنسبة لنوع الجمل التى تحتويها، وهذا يؤكد مرة أخرى على أن طبيعة الإنتاج الفكرى تتجاوز الاختلافات فى كل من الشكل والقائمين بالاستخلاص.

٣-٣-٢ تأثيرات ذلك على الاسترجاع

**** المقومات التركيبية Syntactic Features**

دراسة طول المستخلصات بصفة عامة وكذلك النسبة بين هذه المستخلصات والوثائق المصدرية هما مقومان تركيبيان لهما تأثيرات خطيرة على البحث على الخط المباشر .. وعلى الرغم من أن أحدا لم يكتشف هذا الجانب من الاستخلاص

إلا أن اختلافات حسابات إجمالى الكلمات قد يكون له صلة بمدى شمول المستخلص وتمثيله للمحتوى الأولى للوثيقة المصدرية.

كما أن طول المستخلص يمكن أن يعكس تخصيص تمثيل الوثيقة أو على الأرجح توفيقا بين الشمول والتخصيص. وقد لاحظ العديد من المنظرين فى علم المعلومات أن هذه العوامل تؤثر بدرجة كبيرة فى عملية الاسترجاع النهائية خصوصا بالنسبة لدقة البحث Search precision واستراتيجية البحث المثالية.

وإذا كان العدد المتوسط والوسيط للكلمات فى كل مستخلص كانت متشابهة داخل كل مجموعة إلا أن العدد الكلى لكلمات المستخلص اختلفت كثيرا فى كل مجموعة وهذه الحقيقة ذات دلالة بالنسبة للاسترجاع على الخط المباشر .. وهذه تبين بالضرورة توزيعات طبيعية normal لحسابات مجموع الكلمات لكل مجموعة ولكنها تشير كذلك إلى تغييرات كبيرة فى المحتوى و(أو) مستوى التفصيل فى المستخلصات كوحدات فى كل عينة.

وأخيرا فلابد من المقارنة بين المستخلصات والوثائق الأصلية قبل أن نصل إلى خاتمة أو استنتاجات وإن كان هذا الجانب يشير إلى نتائج بحث غير موثوق بها ..

ومن الملاحظات التى يجب إبرازها أن الفروق فى طول الكلمات الكلى يشير إلى أن المستخلصات التاريخية (للمقالات والرسالات) لا تمثل وثائقها المصدرية بنفس المستوى المعلوماتى كما وجدت فى مستخلصات العلوم.

وعلى الرغم من أن تقرير نسبة الضغط المناسبة optimal compression ratio يجب أن يعتمد على عوامل عديدة مثل تجانس الوثيقة واحتياجات المستفيدين من المستخلص إلا أننا يمكن أن نستنتج على الأقل، أن مستخلصات مقالات التاريخ قد زودتنا بمصطلحات بحثية أقل كثيرا لقواعد البيانات على الخط المباشر وكذلك زودتنا بمواد أقل يمكن أن نبنى عليها أحكام الصلاحية relevance .

هذا ويمكن ربط الفروق بين المجموعات بالنسبة لطول المستخلص بعوامل متعددة .. فهناك الافتراض المنطقى الذى يشير إلى أنه الوثيقة الأطول تتطلب كلمات وجمل أكثر من الوثيقة الأقصر .. ويؤيد هذه الفكرة أن المستخلصات فى DAI كانت أكثر من ضعف طول مستخلصات الكيمياء وعلم النفس وكانت أطول

بمقدار أربع مرات من مستخلصات مقالات التاريخ .. ولكن ذلك تناقض مع مستخلصات مقالات التاريخ، فعلى الرغم من أن مقالات التاريخ شكلت ثانى أطول مجموعات الوثائق، إلا أن مستخلصاتها احتوت على أقصر مجموعة بدائل surrogates.

وهناك تفسيران مناسبان لهذا الوضع أولهما أن طبيعة مقالات التاريخ قد تكون مسئولة عن هذه الظاهرة، أى أنه من الممكن لكلمات أقل أن تمثل المقالات التاريخية وليس الأمر كذلك بالنسبة لتقارير البحوث فى العلوم .. ولكن الاحتمال الثانى أو التفسير الثانى هو أن مستخلصات مقالات التاريخ لا تحتوى على كثافة معلوماتية كما هو الحال بالنسبة لبدائل المقالات العلمية.

وهذا الأمر متوقع بالنسبة للاستجابة لرغبة ABC- Clio's فى التحكم والاستخلاص وتكاليف اختزان البيانات وبالتالى فالتعليمات الموجهة إلى المستخلصين هو أن يتراوح التلخيص من (٥٠) إلى (١٠٠) كلمة.

وهناك عامل تركيبى هام آخر بالنسبة للاسترجاع على الخط المباشر وهو طول الجملة، وليس من المستغرب أن المستخلصات تعكس ممارسة الكتابة الإنسانية الشائعة بالنسبة لاختلاف طول الجملة، فطول الجملة له علاقة كبيرة بالنسبة لاستخدام نظام الديالوج DIALOG حيث يعطى تعليمات للحاسب الآلى للبحث عن عدد مرات حدوث مصطلحات محددة داخل نفس الجملة، وهذا الموضوع ككل يفتح لنا أبواب البحث المستقبلى الذى يتركز حول الخصائص التركيبية كطول الجملة والمستخلص وعلاقتهما بالمحتوى الدلالى والأداء الاسترجاعى.

هذا وعدد كلمات العنوان لكل عينة يزودنا بالدليل عن نجاح التكشيف والبحث على الخط المباشر فى العلوم أكثر منه فى الإنسانيات فقد احتوت عناوين مقالات الكيمياء على ٢٨,٥٪ من كلمات العنوان الدالة بالنسبة لجميع المجموعات .. كما يمكن للباحث أن يستنتج من البيانات الواردة بالبحث نسبة كلمات العنوان إلى كلمات الوثيقة قد كانت أعلى فى مجال الكيمياء من غيرها من المجالات فقد وصلت هذه النسبة إلى ٠,٠٠٤ بينما وصلت فى علم النفس إلى ٠,٠٠٢ وفى مقالات التاريخ إلى ٠,٠٠٠١ وفى رسالات التاريخ إلى ٠,٠٠٠١ .

ونظرا لأن العديد من مصطلحات الكشاف تستمد من عناوين الوثائق، وفى حالة كل من كشافات كويك وكوك Kwic / Kwoc فهى تأتى جميعها من هذا المصدر، ونظرا لأن العديد من البحث على الخط المباشر يشمل حقل العنوان بالإضافة إلى مصطلحات الكشاف، فيمكن أن نستنتج أن العدد الأكبر من الكلمات التى تحتويها عناوين مستخلصات الكيمياء يمكن أن يؤدى إلى بحث أكثر فاعلية.

**** المحتوى الدلالى Semantic Content**

هناك أربعة استنتاجات بالنسبة للبحث على الخط المباشر تأتينا من هذه الدراسة، وأولها أنه يمكن للباحثين فى المجالات العلمية استخدام معايير (ANSI/ISO) كمرشد لهم لأنواع المحتوى المحتمل وجوده فى مستخلصات قواعد البيانات ثانيها: أنه يجب على الباحثين أن يكونوا على دراية ووعى بأن تمثيل فئات محتويات (ANSI/ISO) ليس منتظما بالنسبة لجميع المستخلصات داخل قاعدة بيانات علمية معينة. وثالثها: أنه يجب على الباحثين فى النص الحر للمستخلصات .. خصوصا وأنه غير واضح حاليا أى العناصر الشائعة فى قواعد البيانات غير العلمية، ولكن الأسماء المحددة والتواريخ والأماكن والعناوين يمكن أن تزودنا بالاسترجاع الأفضل فى حقول النصوص الحرة. وأخيرا فيجب أن يكون الباحثون فى قواعد البيانات غير العلمية واعين بالتمثيل غير المتوازن لهذه الوثائق فى ملفات قواعد البيانات.

نتائج عامة فى هذه الدراسة:

(١) تعتبر المستخلصات الإعلامية أهم عناصر نقاط الإتاحة فى استرجاع المعلومات.

(٢) الإتاحة الموضوعية فى الإنسانيات ليست عسيرة نظراً لأن أكثر المصطلحات استخداما هى أسماء الأعلام المفردة وبالتالى يمكن إنشاء نظم استرجاع فعالة (وإن كانت هناك نسبة أقل من المصطلحات الأقل دقة حيث يتميز تعريفها بالتغيير عبر الزمن أى أن لها مدى واسعا من المعانى).

(٣) الاسترجاع الأفضل فى مجالات الإنسانيات يستدعى إثراء العناوين بمصطلحات أخرى لزيادة الاسترجاع.

(٤) بمقارنة درجة حداثة مراجع الدوريات الجارية فى كل من العلوم الطبيعية والاجتماعية، تبين أن الوسيط لمنتصف الحياة يصل إلى حوالى ست سنوات فى الإنتاج الفكرى لكل منهما.

(٥) تمثيل فئات محتويات المعايير ليس منتظما بالنسبة لجميع المستخلصات داخل قاعدة بيانات علمية معينة.

(٦) هناك شك فى إمكانية كفاية المعايير الدولية لاستخلاص جميع أنواع الإنتاج الفكرى.

** توجيهات وتوصيات للبحوث المستُقبلية:

لقد اعترف علماء العلوم الطبيعية منذ زمن بعيد بأهمية وفائدة المستخلصات وكانت هذه المستخلصات ضمن معظم المقالات المنشورة وضمن الكشافات الموضوعية أما بالنسبة للمشتغلين بالإنسانيات فقد كانوا أكثر بطئا من نظرائهم العلماء فى الاهتمام بالمستخلصات على الرغم من أهمية بدائل النص المختصرة بالنسبة لاسترجاع الوثائق فى المجالات الإنسانية كالتاريخ والأدب.. خصوصا والإنتاج الفكرى فى مجال الإنسانيات يفتقد فى أحيان كثيرة للعناوين الإعلامية الدالة على المحتوى كما تمثل المقالات فى المجالات الإنسانية مشكلات تكشيف خطيرة.

وعلى كل حال فالحاجة إلى بدائل كافية واضحة نظرا للأبعاد المختلفة للأرشيف فى الإنسانيات وسرعة تلف محتويات العديد من المكتبات وتزايد عدد المكتبات البحثية التى تدخل فى مشروعات تعاونية لبناء المجموعات .. أى أن المشتغلين بالإنسانيات سيجدون مع تزايد الإنتاج الفكرى الحاجة الماسة لاستشارة نظم الاسترجاع الببليوجرافى وكذلك ضرورة توفر بدائل الوثائق عند عدم توفر النصوص الكاملة.

وتخدم الدراسة التى بين أيدينا غرضين أولهما أن الباحثين على الخط المباشر سيجدون المستخلصات المتوفرة حاليا فى كل من الكيمياء وعلم النفس والتاريخ

ذات قائدة عند البحث فى قواعد البيانات لهذه المجالات وثانيهما أنه مجهود بحثى للحث على إنتاج مستخلصات إعلامية ومنتظمة فى المجالات غير العلمية.

ونظراً لأن محتوى المستخلصات التى أعدها كل من المؤلف أو المستخلص فى مجال التاريخ مختلفة تماما عن المستخلصات العلمية، ونظراً لأن المستخلصات التاريخية لها علاقة ضعيفة بمعايير (ANSI/ISO) فإن هذه الدراسة التى بين أيدينا يمكن أن تعتبر كقاعدة لمزيد من تحليل مستخلصات الإنسانيات بما فى ذلك التاريخ.

كما أن الدراسة التى بين أيدينا تشكك فى إمكانية وجود معايير عالمية تنسحب على جميع أنواع الإنتاج الفكرى ولجميع المجالات العلمية .. كما أن هناك ضرورة لوضع معايير البدائل داخل كل مجال علمى على حدة.

الملحـــق

تعليمات التكويد الخاصة بتحليل وظائف الجمل التى تحتويها المستخلصات الخاصة ببحوث مختلف المجالات العلمية

يرجى توضيح وظيفة كل جملة فى المستخلصات المرفقة وذلك طبقا للخطة الموضوعة أدناه، وقد اعتمدت هذه الفئات على المعيار الوطنى الأمريكى لكتابة المستخلصات وهو ANZI Z 39.14 - 1979 .

(١) قم بوضع واحد أو أكثر من رموز التكويد على الخط المقابل الخاص لتصنيف الجمل، يرجى استيعاب البدائل بعناية بما فى ذلك البند الخاص بالاختيار "ليس مما هو مبين" NOTA، ولا تحاول إقحام جملة فى فئة معينة.

(٢) قم بشرح الاختيار الذى تفضله عن طريق
أ) التأشير على كلمة مفتاح فى النص أو
ب) كتابة الشرح الذى تريده فى عمود "السبب" Reason (كلمة أو كلمتين كافيتين).

(٣) عند اختيارك لفئة "ليس مما هو مبين" NOTA يرجى اقتراح فئة بديلة تراها تشرح وظيفة تلك الجملة داخل النص.

الخلفية : B = Background

تشمل الجمل ذات معلومات الخلفيات الجمل التى تحيل إلى البحوث السابقة و(أو) الحقائق والنتائج والأوصاف والنظريات ... إلخ المعروفة قبل القيام بهذه الدراسة.

الهدف / النطاق : P/S/ =Purpose / Scope

تشمل الجمل المتصلة بهدف الدراسة / أو ورقة البحث ومتضمنة الأهداف الأولية ونطاق الدراسة أو الأسباب التى من أجلها تم كتابة الوثيقة.

الفروض : H = Hypotheses

الجمل التى تشمل الفروض التى يتم اختبارها فى الدراسة هى هذه الفئة، وهذه تختلف عن الاستنتاجات Conclusions والتى يمكن أن تناقش سبب قبول أو رفض بعض الفروض.

المنهجية M = Methodology

تشمل الجمل الدالة على المنهجية تلك التى تحدد الأساليب الفنية ونصف المبادئ المنهجية ومدى العلميات والدقة التى تكون فى الدراسة، أما بالنسبة للوثائق التى تهتم بغير العمل التجريبى، فإن بيانات المنهجية نصف مصادر البيانات وكيفية تطويع البيانات.

النتائج: R = Results

نصف بيانات النتائج ما توصل إليه الباحث نظريا وتجريبيا، وكذلك البيانات المجمعة والعلاقات والارتباطات الملاحظة، وكذلك التأثيرات وحدود الدقة والثقة والصحة. accuracy/reliability/ validity

الاستنتاجات : C = Conclusions

الجمل التى تشمل الاستنتاجات يمكن أن تصف تأثيرات النتائج خصوصا بالنسبة لعلاقة هذه النتائج بهدف البحث أو سبب إعداد الوثيقة، ويمكن أن تربط الاستنتاجات كذلك بالتوصيات والتقييمات والاقتراحات والعلاقات الجديدة والفروض المقبولة أو المرفوضة.

ليس مما هو مبين أعلاه: N = None of the above

إذا لم تكن أى واحدة من الفئات أعلاه تمثل وظيفة الجملة داخل المستخلص فإن الرمز (N) يجب استخدامه للدلالة على هذه الجملة، وهذه يجب أن يتبعها جملة قصيرة تصف وظيفة هذه الجملة أو نوع المحتوى وذلك فى عامود "الأسباب" Reasons .. تعليقاتك تعتبر ذات أهمية نظرا لأنها ستقدم بصيرتك ووجهة نظرك أبعد مما هو موجود فى معايير ANSI.

المحتوى المختلط MC = Mixed Content

قد تحتوى بعض الجمل ـ خصوصا الجمل الطويلة أو المعقدة ـ على معلومات تخدم أكثر من وظيفة واحدة داخل المستخلصات، وعلى سبيل المثال فقد تحتوى جملة على عبارة (أو عبارات) أو فقرة (أو فقرات) شاملة للنتائج Results بينما يقدم جزءاً آخر استنتاجات conclusions .

الفصل الثامن

الاستخلاص الآلى: أضواء من الإنتاج الفكرى الأجنبى خلال أربعين عاما

مقدمة

ارتبط الاستخلاص الآلى بالتكشيف الآلى بعد أن يسر الحاسب الآلى للمهنيين فى المعلومات مهام معالجة وتجهيز البيانات، وإن كان الأمر بالنسبة للإستخلاص الآلى أكثر صعوبة نظرا لعدم توفر القواعد السليمة اللازمة لبرمجة الحاسب الآلى للقيام بالأعمال الفكرية المتصلة بالاستخلاص..

وسنتناول هذه الدراسة المحاور السنة التالية:

أولاً : الاستخلاص الآلى: هل هو الاقتباس الآلى؟

ثانياً : إجراءات الاستخلاص الآلى وبعض مشكلاته.

ثالثاً : أضواء على بحوث التلخيص الآلى Summarization

رابعاً: نظم معاصرة وبحوث مستقبلية فى الاستخلاص الآلى.

خامساً: الذكاء الاصطناعى والنظم الخبيرة فى حقل استرجاع المعلومات

سادساً: بعض النتائج والتوصيات.

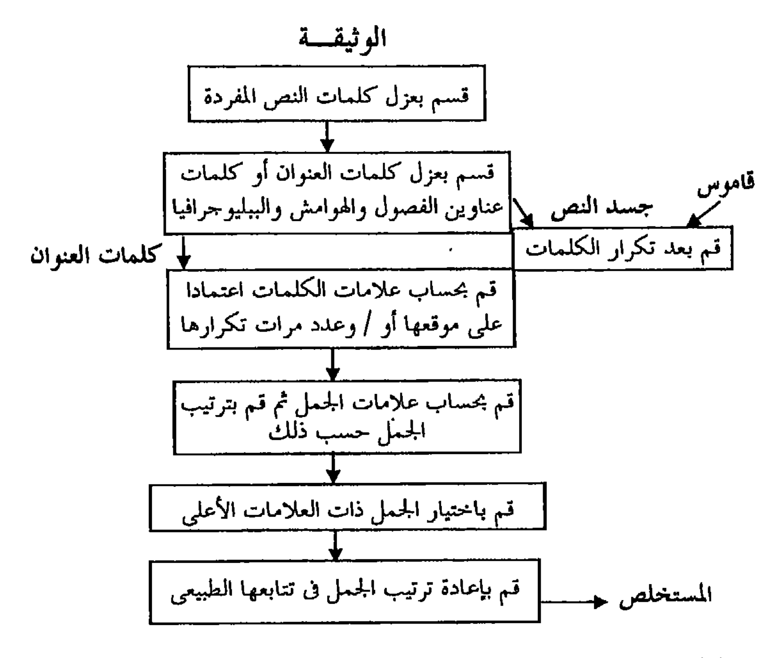
أولاً : الاستخلاص الآلى: هل هو الاقتباس الآلى؟

(١) خطوات نظام اقتباس الجمل المثالى:

يمكن أن يقال بأن طرق التجهيز الآلية لمعالجة النص المعتمدة على تحديد أهمية المصطلح والجملة لم تستخدم فقط فى التكشيف ولكنها استخدمت كذلك فى أغراض الاستخلاص الآلى، وفى الواقع فإن الإجراءات الآلية تتضمن عملية اقتباس حيث يعرف المستخلص ببساطة بأنه مجموعة صغيرة من الجمل المقتبسة من الأصل والتى يفترض أنها هامة بالنسبة لتمثيل المحتوى، وتبدأ طرق الاقتباس المستخدمة عبر السنين بحساب دلالة الكلمة أو الجملة بطريقة مشابهة لحساب وزن المصطلحات فى التكشيف الآلى.

(Quoted by Salton, G., 1975: 107 from Edmundson, 1961, 1964, 1969)

وفيما يلى خطوات نظام اقتباس الجمل المثالى (Edmundson, 1964)



فعملية الاستخلاص تسهل الحصول على المعرفة حيث تستبعد العرضى وتحفظ الضرورى من المعلومات، وقد أشار الباحث وينوجراد (inograd, 1984) إلى أن كفاءة الاستخلاص تتضمن القدرة على شرح النقاط الرئيسية للوثيقة بطريقـــة مختصرة.

(٢) عملية الاستخلاص العامة والاتجاه نحو المكينة:

(GAP - General Abstracting process)

وهذه تتضمن سلسلة من العمليات وتشمل التحول الذى لابد أن ينال الوثائق النصية من تركيبها السطحى إلى وصف البناء (أو المحتوى) المتعمق، وبالتالى فالمستخلص يتضمن عمليات شاملة تفسيرية انتقائية ثم تركيبية، الغرض منها إعادة تركيب المعلومات النصية كوثيقة جديدة ممثلة للوثيقة على نطاق أو مقياس مصغر.

أما الباحثان كريمنز وراولى (Cremmins, 1982 and Rowley, 1988) فقد بسطا هذه العمليات واقترحا طريقة عملية معتمدة على الأفكار الكونية، فكريمنز Cremmins يشير إلى النموذج بهذه الطريقة كما يلى "يتم أداء العمليات العقلية فى مراحل أربع بالتقريب وهى:

(أ) التركيز على المظاهر الأساسية للمواد المراد استخلاصها.

(ب) تحديد المعلومات ذات العلاقة.

(ج) الاقتباس والتنظيم وتقليل أو تصغير المعلومات المتعلقة إلى وحدة متماسكة، عادة فقرة مطولة.

(د) تنقية المستخلص عن طريق التحرير.

أما الباحث راولى Rowley فقد ذكر خمس خطوات وهى:

أ) اقرأ الوثيقة لفهم محتوياتها والتعرف على نطاقها.

ب) اكتب ملاحظاتك عن النقاط الأساسية فى الوثيقة.

ج) اكتب مسودة مستخلص من الملاحظات المدونة.

د) قم بمراجعة مسودة المستخلص للترقيم والهجاء والدقة والحذف والتحديد.

هـ) اكتب المستخلص النهائى.

وإذا كان هذا الذى سبق يتصل بتعريف المستخلص وطريقة إعداده، فالباحث مولينا (Molina, M., 1995, p. 226) يذهب إلى أن هناك اتجاهات فى السنوات الأخيرة نحو ميكنة عمليات الاستخلاص ولكن النتائج ليست مرضية تماماً، ويعزز هذا الرأى ما ذهب إليه الباحث بيس (Paice, 1990, p. 175) إلى أنه من السهل التعرف على عيوب الاقتباسات المنتجة بواسطة الحاسب الآلى والتى تتمثل فى جانبين أولهما عدم توافر التماسك Cohesion وعدم توافر التوازن Balance، وانتهى فى دراسته إلى أنه يبدو أن التقدم فى توليد المستخلصات الآلية يجب أن يعتمد على وجود نظرية مرضية لبناء النص.

وأخيراً فيشير الباحث هاتشنز (Hutchins, J., 1987) إلى أن هناك أسباباً قوية تدعونا لمزيد من الدراسات الجادة لعمليات التلخيص ويشترك فيها علماء المعلومات واللغويون والباحثون فى الذكاء الاصطناعى وكثير غيرهم، على أن

يكون الهدف الأول ليس "الأتمتة" ولكن الفهم الأساسى. فالتلخيص ضرورى لفهم النص ولإنتاج النص وهو حاسم لتنظيم المعلومات والمعرفة ولكننا نجهل كل جانب من جوانب هذه العمليات.

ووضع هاتشنز نموذجا إجرائيا للاستخلاص له الدقة والثقة المرغوبة، حيث يمكن لهذا النموذج أن يقدم لعلم المعلومات الشرح الواسع العملى لتمثيل الوثيقة بالنسبة لمحتواها.

ثانياً: إجراءات الاستخلاص الآلى وبعض مشكلاته:

إذا كان الاستخلاص اليدوى مرتبط وظيفياً بالتكشيف اليدوى، فالاستخلاص الآلى مرتبط بالتكشيف الآلى، وقد حمل الباحث لوهن Luhn مثلاً أفكاره المتصلة بتكرار الكلمات من التكشيف إلى الاستخلاص وقام بوضع الأساليب التى أدت إلى أولويات الترتيب Priority Rating لكل جملة فى النص اعتمادا على درجة تكرار حدوث الكلمات، وتم اختيار وترتيب الجمل حسب درجة تحميلها بالكلمات العالية الدلالة كجزء من المستخلص الآلى.

لقد اعتمدت قاعدة لوهن Luhn الخاصة بتحديد الأولويات لا على مجرد وجود الكلمات فى الجملة، ولكن على علاقات الكلمات المفتاحية داخل الجملة، وتضمنت هذه الطريقة نظرة شاملة على الجملة لتقسيمها إلى مقاطع phrases بناء على الكلمات الدالة، ثم ترشح هذه الأجزاء للتجهيز إذا لم يكن هناك أكثر من خمس كلمات غير دالة من بين الكلمات الدالة، ويلى ذلك القيام بحساب الأولوية بالنسبة للجملة عن طريق جدولة الكلمات الدالة فى كل فقرة على حدة ثم مضاعفة هذه القيمة وقسمتها على العدد الكلى لكلمات الفقرة، وقد اتبع لوهن (Luhn, 1958) الإجراءات التالية:

١- قائمة استبعاد جميع الكلمات غير ذات الدلالة من المعالجة.

٢- درجة حدوث جميع الكلمات الباقية يتم عدّها، ثم ترتب طبقاً لدرجة تكرار حدوثها (ويمكن استخدام الجذور بدلاً من الكلمات).

٣- جميع الكلمات التى تحدث أكثر من عدد (س) من المرات، يمكن تعريفها بأنها ذات تكرار عال أو كلمات دالة.

٤- يتم تحديد الجمل التى تحتوى على تركيز لهذه الكلمات ذات التكرار العالى، وفى هذه الحالة يمكن اعتبار كلمتين متعلقتين ببعضهما داخل إحدى الجمل إذا لم يكن هناك بينهما أكثر من أربع كلمات.

٥- يتم حساب عامل الدلالة لكل جملة كما يلى:

(أ) حساب عدد التجمعات Clusters فى الجملة (والتجمع هو أطول تجمع للكلمات المرتبطة ببعضها بكلمات ذات دلالة، وحيث لا يكون بين الكلمات ذات الدلالة هذه أكثر من أربع كلمات أخرى).

(ب) يحسب عدد الكلمات ذات الدلالة فى التجمع، ثم يتم قسمة مربع هذا العدد على العدد الكلى للكلمات داخل التجمع.

(ج) يمكن أن يعرف عامل الدلالة للجملة إما بقيمة أعلى تجمع أو جمع قيم جميع التجمعات فى الجملة.

ولعل هذا العرض السابق أن يكون أكثر تعقيداً فى شرحه منه فى تطبيقه العملى، ويمكن شرحه بسهولة من خلال المثال التالى:

ABCD* EF* G* HIJ* KLMNOPQR.

فإذا اعتبرنا كل حرف من هذه الحروف ممثلاً لكلمة، وأن الكلمات التى عليها علامة هى الكلمات التى نعتبرها ذات دلالة، فالتجمع الذى يتكون بالكلمتين D,J يحتوى على أربع كلمات ذات دلالة، وبالتالى فإن عامل الدلالة لهذا التجمع هو $\frac{4^2}{7} = 2.3$ أو ٢,٣ وهذا هو عامل الدلالة للجملة نظراً لأنها تحتوى فقط على تجمع واحد.

وطبقاً لإجراءات Luhn فإن الكلمات التى تحتوى على أعلى عوامل الدلالة هى التى يتم اختيارها وطباعتها، وذلك فى التتابع الذى يحدث فيه بالنص، وذلك لتكوين المستخلص، وينتهى المستخلص عادة اعتماداً على عدد محدد من الجمل أو اعتمادا على عدد الجمل المطلوبة للوصول إلى نسبة مئوية معينة من نص الوثيقة الكلى.

وعند معالجة الوثائق الطويلة جداً، فمن المفضل أن تقوم البرامج باختيار وطباعة الكلمات ذات الدلالة لكل جزء من المطبوع، ويمكن وضع بعض المعايير والموازين لضمان وجود الجمل التى تحتوى على واحد أو أكثر من التكرارات المطلوبة من الكلمات، وأن هذه الكلمات سيتم اختيارها لتكون ضمن المستخلص.

ومن الواضح أن مثل هذه المستخلصات التى تتكون بهذه الطريقة سوف لا تشبه المستخلصات المعدة بواسطة الإنسان، وذلك لأن بعض الجمل قد تأتى من الفقرة الأولى، وقد تأتى من آخر الفقرات أو من وسط العمل أو غير ذلك، أى أن الاقتباس قد يظهر أنه غير متعلق ببعضه، وليس ذلك شيئا هاما للغاية إذا ما استطاعت الجملة المختارة أن تقدم لنا صورة دقيقة لمحتوى الوثيقة.

وقد نجد بعض الباحثين الذين لا يتفقون مع هذا الرأى، ويصرون على أن الاقتباسات الآلية يجب أن تكون أكثر تماسكاً واستمرارية. (Mathis, et al, 1973).

وهناك بعض الباحثين مثل راش (Rush, 1971 ... cited by Lancaster, F., 1991 p. 237). وماتيس من الذين يرون أن المدخل الصحيح للاقتباس يجب أن يتضمن معايير لرفض الكلمات، فضلاً عن إمكانية اختيارها، ويرى راش وزملاؤه أن تقييم الجمل يجب أن يأخذ فى اعتباره التأثير النصى، ويمكن أن يكون ذلك مفتاحاً لقبول أو رفض الجملة، وبالتالى فإن طريقة الاقتباس تعتمد على مضاهاة النص أمام قائمة ضبط الكلمات Word controlled list وهذه القائمة تشمل تعبيرات إذا وجدت فى الجملة، فإن ذلك يؤدى إلى رفضها بينما يتم قبول الجملة واختيارها إذا تضمنت عددا قليلا من هذه التعبيرات، هذا والجمل التى تحتوى على كلمات ذات دلالة من العنوان يمكن أن يتم اختيارها أيضا، أما بالنسبة لمعيار التكرار فهو يستخدم فقط لتعديل الأوزان المرتبطة بالمفاتيح الإيجابية أو السلبية فى قائمة ضبط الكلمات.

هذا وقد أشار العالم "بيس" (Paice, 1981 ... cited by Lancaster, F., 1991, p. 240) إلى أن الاقتباس الآلى المعتمد على التعرف على الكلمات، هو أحد الأدلة الهامة لما تحتويه الوثيقة.

وهناك مدخل آخر يتمثل فى التركيز على بعض جوانب أسلوب الوثيقة، فعلى سبيل المثال يمكن للحاسب الآلى أن يبرمج لاصطياد الجملة الموضوعية Topical sentence كالأجزاء الأولى أو الأخيرة من الفقرات، وهناك بعض المحاولات التى تمت بالنسبة لإنشاء مكنز يمكن اختزانه فى الحاسب الآلى للمعاونة فى تحديد الكلمات والتعرف عليها لاقتباسها، ومما سبق يتضح أن مجال الاستخلاص الآلى لم يحرز حتى الآن إلا نجاحاً جزئياً، وهو يمدنا فى بعض الأحيان بمعلومات مختصرة موجزة مفهومة، وفى أغلب الأحيان تظهر المستخلصات كخيوط من الجمل التى لا رابط بينها (Cleveland, D., 1990, p. 236).

ثالثاً: أضواء على الجهود الحالية والمستقبلية لبحوث التلخيص الآلى:

Automatic Summarizing

(١) مقدمة:

عقدت حلقة دراسية فى مدينة داجستول Dagstuhl بألمانيا فى ديسمبر ١٩٩٣ تحت عنوان "تلخيص النص من أجل الاتصال الذكى" (Endres-Niggemeyer, B. et al, 1993) وقد استهدفت الحلقة نظرة أكثر شمولاً للمجال بحيث تضم تلك النظرة المتطلبات والمصادر والإمكانيات اللازمة للتلخيص Summarizing وذلك لتحقيق غرضين أولهما تقديم دافع أفضل للتطبيقات متوسطة المدى وثانيهما لتقديم خلفية صلبة للبحوث على المدى الطويل، هذا فضلاً عن رؤية التلخيص رؤية سليمة، ليس بالنسبة لدورها التقليدى فى تقديم بدائل ثابتة fixed Surragates للوثائق المختزنة، ولكن كنشاط ديناميكى لإنشاء ملخصات تلائم احتياجات المستفيدين.

٢) مجالات المصادر اللازمة للتلخيص الآلى:

يعتمد العمل المستقبلى فى التلخيص الآلى على مجالات متعددة منها التلخيص الإنسانى والجانب اللغوى لطبيعة الحديث discourse والنصوص Texts وعلى الدراسات النفسية الخاصة بتجهيز ومعالجة الحديث والحوار وعلى البحوث المتصلة بملاءمة النظم التفاعلية للمستفيدين منها وأخيراً على تجهيز اللغة الطبيعية بصفة عامة.

فبالنسبة **للتلخيص الإنسانى** كنشاط مهنى فإن له جوانبه العملية فضلا عن الخطوط المرشدة التى يمكن أن تكون مفيدة، وقد تم البحث فى هذه الجوانب كمصدر من مصادر نماذج التلخيص لا سيما بالنسبة للنماذج الإجرائية.

أما **العمل اللغوى للحديث** فقد حاولت الدراسات التعرف على هذه الظاهرة بالنسبة للحوار المسموع أو النص المكتوب، واقترحت بعض نماذج بناء النص والتى تبدو ذات علاقة بالتلخيص وهذه تشمل الفئات الوصفية والوظيفية لوصف وحدات النص بأحجامها المختلفة. أما البحوث **المتصلة بتقسيمات النص** Text typology فلها أهميتها بالنسبة للتلخيص نظراً لأن بعض الأنواع geners لا تحفزنا فقط لتبنى استراتيجيات التلخيص بل هى تتطلب ذلك لكل نوع.

أما بالنسبة **للدراسات السيكولوجية** (النفسية) الخاصة بالقراءة أو الاحتفاظ بالحديث discourse فى الذاكرة كما تظهر فى التلخيص، هذه الدراسات يمكن أن تلقى الضوء على معالم النص وأجزائه التى يمكن تذكرها أو على خصائص النص التى تخدم فى التعرف على أجزائه الهامة التى يتم تذكرها.

(٣) صياغة الإطار modelling Framework

إذا قمنا بتعريف الملخص Summary على أنه الاشتقاق المكثف للمصدر، أى تقليص المحتوى بالاختيار أو بتحديد ما هو مهم، فإن هذا التوصيف لا يأخذنا بعيدا، ولكننا نستطيع تحليل مكونات التلخيص بحيث يكون لدينا فى البداية مجموعة العوامل التى تؤثر على التلخيص ثم يكون لدينا بعد ذلك نظرة لعملية التلخيص الأساسية.

وإذا أخذنا هذا كله كإطار فيمكننا حينئذ أن نختبر الخصائص المميزة لأى متطلبات تلخيص محددة، وبناء على ذلك فهناك ثلاث مجموعات من العوامل التى تؤثر على التلخيص وهى: عوامل مدخلات Input factors والتى تحدد خصائص المادة التى سيتم تلخيصها، وهناك عوامل الغايات Purpose factors وهذه تحدد متطلبات التلخيص ثم هناك كذلك عوامل المخرجات Output factors والتى ستحدد بالمدخلات والغايات، ولكنها ستتضمن كذلك تشخيص التلخيص المنتج.

وعوامل المدخلات يمكن أن تجزأ إلى فئات Categories تحت الشكل والموضوع والوحدة، والشكل يتضمن المقياس والتنظيم الواضح، أما الموضوع فيشير إلى درجة التخصص فضلاً عن حقل التخصص أما الوحدة فقد تضم واحدة أو مجموعة مواد.

أما **عوامل الغايات:** فتشمل الحالة أو وضع المادة والسياق الإجرائى لاستخدام الملخص أى الوظيفة المحددة للملخص.

أما **عوامل المخرجات:** فتشمل فئة المادة أو شكلها Format وهذه تغطى الإخراج والأسلوب. ومن هذه العوامل يمكن إعداد النموذج العام الذى يتضح فيه التمييز بين تفسير النص المصدرى وبين توليد النص الملخص.

ويعتبر النموذج ذات أهمية وقيمة خاصة من الناحية البحثية لأنه سيتطابق مع وجهات نظر مختلفة لتمثيل المصدر أو الملخص وتوضيح المسافة بين النص أو الحديث وتمثيله، كما يساعد النموذج فى تأكيد النقاط ذات الدلالة، كما يسهل النموذج دراسة تأثيرات صفات العوامل المختلفة على التلخيص وتطبيقاته.

(٤) الإسهامات ذات الارتباط بالتلخيص:

تهتم الدراسات والبحوث المتصلة بالتلخيص الآلى حتى وقتنا الحاضر بالمعالجة الإحصائية للمفاتيح الخاصة الممثلة للمحتويات الهامة للوثيقة، مع إهمال العديد من عناصر نصوص المصدر، فالدراسات الخاصة بالتلخيص الإنسانى لها ارتباط بالصفات التنظيمية والمعجمية Lexical للنص الخاص بالمصدر وبالتالى لها أهمية بالنسبة للتلخيص، أما لغويات النص Text Linguistics فتركز بصفة أساسية على الأنواع المختلفة للنصوص بما فى ذلك نصوص الملخصات كنصوص قائمة بذاتها وليس على عمليات التحول والتى يعتبر التلخيص جزءاً أساسياً منها، أما تجهيز الحديث discourse processing فهو يركز على الصفات العامة generic للحديث ودرجة ترابطه وتركيزه وكذلك على مراحل التفسير، أما البحوث المتصلة بالملاءمة للمستفيدين User adaptation فهى موجهة اساساً نحو مرحلة التخليق Synthesis وانتاجية النص أى أنها تتناول كلاً من عوامل الغايات والمخرجات السابق الإشارة إليها، وأخيراً فالبحث فى تجهيز ومعالجة اللغة

الطبيعية فهو يتوسل بالإجراءات المحسبة سواء بالنسبة لتفسير الجملة أو الحديث أو الإجراءات المحسبة لتوليد النص.

(٥) التوجيهات المستقبلية:

لقد وضعت الحلقة الدراسية فى داجستول Dagstuhl أربعة خطوط للعمل المستقبلى فى بحوث وتنمية التلخيص الآلى وهى:

أ - هناك دراسات مطلوبة عن الأشكال التى يمكن أن تكون عليها الملخصات وعن الطريقة التى تستخدم بها.

ب- ملاحظة كيفية القيام بالتلخيص الإنسانى يمكن أن يفيد فى تصميم الطرق المحسبة.

ج- لابد من تطوير استراتيجيات التقييم الكافية.

د- يمكن تقدم الاستراتيجيات المحسبة عن طريق اختبار الطرق القائمة مع وضع وتطوير اتجاهات جديدة على المدى البعيد.

(٦) بعض أوراق البحوث التى قدمت للحلقة الدراسية:

تقع أوراق البحوث المقدمة فى ثلاث مجموعات، تتوجه **المجموعة الأولى** حول التلخيص الإنسانى وأن تحليل المهارات والعمليات التى يقوم بها المستخلصون الخبراء من شأنه المعاونة فى أداء التلخيص الآلى. أما **المجموعة الثانية** من أوراق البحوث فتصف النظم التى يتناول بعضها العملية كلها من نص المصدر إلى نص الملخص ويتناول بعضها توليد ملخصات المخرجات. أما **المجموعة الثالثة** فتتناول صفات المستخلصات حيث تبين الطريقة التى تعتمد على المجالات المعرفية الأخرى مثل لغويات النص.

وعلى ذلك فالبحوث المختلفة فى المجموعات الثلاث تزودنا بمدى واسع من الأمثلة داخل الإطار الذى تم طرحه وصياغته من قبل.

ومن أمثلة **أوراق بحوث المجموعة الأولى** ما قدمه أندرى نيجامابر (Endres - Nigge - meyer et al, 1995) من نموذج محاكاة Simulation للمستخلص الإنسانى الخبير، حيث يركز هذا النموذج على أنشطة القائم بالاستخلاص كما

تعكسه استراتيجيات محددة مستخدمة، فضلا عن خطوات التجهيز والمعالجة المتبعة، وهذه الجوانب يمكن ربطها بنص المدخلات من خلال الطريقة التى يمكن تفسيرها باستخدام نظرية البناء البلاغى Rhetorical Structure Theory وفى نهاية مناقشة ورقة البحث يشعر أصحابها بأن نموذج المحاكاة للتخليص المهنى أمر ممكن.

أما مجموعة **أوراق البحوث الثانية** فتصف نظم قائمة فعلاً (Brandow, R. et al, 1995) ونظام براندو وزملاؤه يسمى "أنيس" ANES ويعتمد على اتجاه عام بسيط يهدف إلى اختيار جمل من المصدر للتلخيص عن طريق استغلال إحصائيات الكلمات، كما تناقش هذه الورقة أيضاً التقييم، حيث تقارن بين مخرجات أنيس ANES بمخرجات طريقة روكبوتوم Rockbottom وهى التى تختار بدايات النص فى المصدر، وظهرت بعض مزايا وعيوب كل من الطريقتين وقام الباحثون بشرح أسباب ذلك، ومن بين تلك الأسباب طبيعة المصدر نفسه، فالمصدر الذى يحتوى على قصص أخبار عامة يميل فى بنائه إلى وجود ملخصات افتتاحية، ومع ذلك فقد ثبت أنه من العسير الحصول على ملخصات مترابطة عن طرق ضم بعض الجمل المتفرقة المختارة، وقد كان الباحثون مقتنعين بالنتائج التى توصلوا إليها باتباع المنهج الإحصائى ذى التوجيه الذاتى Statistical/ heuristic method.

أما المجموعة الثالثة لأوراق البحوث فتركز على توليد مخرجات نصوص ملخصة من مدخلات بيانات غير لغوية (Mckeown et al, 1995) حيث يتناول ماكوين وزملاؤه الجوانب اللغوية اللازمة لتقديم معلومات مكثفة فى الملخص فضلاً عن تقديم هذه المعلومات موجزة. والنظامان اللذان قام الباحثون بوصفهما هما ستريك وبلاندوك STREAK & PLANDOC والاثنان يعالجان مشكلة التقديم، ويعتمد الباحثون بشدة على ميدان المعرفة Domain Knowledge وعلى ترتيب العلامات marking التى تميز درجة أهمية المعلومات وذلك للوصول إلى مرحلة تكشيف المحتوى ويلاحظ هنا أن ورقة البحوث هذه (Brandow et al) تستخدم مدخل معرفة وليس مدخلاً معتمدا على النص فضلاً عن أنها أيضاً تستخدم ميداناً محدداً domain - specific وليس ميداناً عاماً general one كما تتناول إعادة صياغة لغوية بدلاً من استنساخ النص.

أما الباحث ميبيرى (Maybury, M.T., 1995) فيتناول بيانات الأحداث مع وضع استراتيجيات لاختيار البيانات وتكشيفها وتقدم المواد ضمن خطة لمخرجات الملخص والتى ستحقق كنص، وتتبع الخطة نحوا عاما general grammer للأفعال الاتصالية والبلاغية لتحديد مختلف أنواع بناء النص، ويهتم الباحث ميبيرى بتصغير البيانات أكثر من اهتمام الباحث السابق ماكوين Mckeown ولكنه يعتمد أقل منه على إرشاد مجال مسبق وهو يستخدم مستخلصين أفراد للحصول على المعلومات المفتاحية من كل من المصادر والملخصات ولكن النتائج لا يمكن أن تؤخذ إلا على أنها ملخصات شارحة indicative أما ورقة البحث الخاصة بالباحثة روثكيجل (Rothkegl, 1995) فهى تعامل المستخلص على أنه نص يتعرض للأحوال العامة والمتخصصة لإنتاج النص، والغاية ـ المحددة بتكوين المستخلص كنص ـ تتحكم فى عملية الاستخلاص كلها، وتصف الباحثة استراتيجيات الإنتاج بالنسبة لعمليات بناء النص والتى يعاد إنشاؤها بنماذج من مجال لغويات النص واللغويات المحسبة text linguistics and computational linguistics.

وهذا يؤدى بدوره إلى علاقة وثيقة بين تمثيل النموذج Model والنص الناتج، أما النموذج نفسه فهو يحتوى على ثلاثة مستويات متكاملة من النص والتى يتم ضمها مع بعضها ثم تمثيلها بناء على أنشطة الكاتب.

ويمكن القول بصفة عامة أن أوراق البحوث فى هذه الحلقة الدراسية تشير إلى أن هناك عملاً محسباً هاماً فى مجال التلخيص، وتتناول البحوث مدى واسعاً من المشكلات المحددة داخل مجال التلخيص وذلك باتباع مناهج ومداخل مختلفة. ومع ذلك فما زال الطريق طويلاً قبل كتابة البرامج القادرة على تفسير النص وليس مجرد توليده، وما زال الطريق بعيداً أمام البناء النصى الكبير مع تحديد محتوى المصدر الهامة باستخدام معلومات الحديث والمجال discourse and domain .

رابعاً : نظم معاصرة وبحوث مستقبلية فى الاستخلاص الآلى:

(١) مقارنة نظام سمارت ونظام دوسكوكس:

لقد تم خلال السنوات الماضية تطوير إجراءات تسمح لنا بوضع بعض الجمل أو الفقرات كتعبير عن طلب للمعلومات ثم مضاهاتها بنص الوثيقة (سواء أكان النص الكامل أو النص المجزء أو بعض أشكال التمثيل). ومن الواضح أن هناك

معايير مختلفة يمكن استخدامها فى وضع أو تحديد درجة التمثيل من النص تعكس درجة مضاهاتها للنص المطلوب، وبالتالى فالنظام الآلى يجب أن يتضمن عدة معايير للمضاهاة الممكنة حتى يختار منها المستفيد.

ولعل أكثر النظم تعقيداً فى هذا الجانب هو نظام SMART الذى وضعه Salton والذى تطور عبر أكثر من ربع قرن من الزمان، وقد صمم نظام سمارت لوضع أوزان رقمية للمواد تعكس درجة مضاهاتها لطلب المستفيد على أن تقدم هذه المواد إلى المستفيد فى ترتيب منطقى، حيث توجد الأوزان العالية فى البداية، وقد ضمن سمارت معايير مضاهاة متنوعة بما في ذلك وزن المصطلحات التى تعكس عدد مرات حدوثها فى قاعدة البيانات، فضلاً عن مضاهاة الجمل ومضاهاة جذور الكلمات، كما يسمح هذا النظام بأن يشمل مكنز يتم وضعه بالمعالجة الإنسانية والمحسبة، هذا وهناك عنصر هام أيضاً فى نظام سمارت وهو صلاحية التغذية المرتدة Relevance feed back ولتوضيح ذلك فإذا كان هناك مخرجات أولية لبحث معين، واستطاع المستفيد أن يتبين أى المواد ذات الصلة وأيها يجب استبعادها، فإن النظام يستطيع أن يعيد حسابات أوزان المواد فى قاعدة البيانات، ويتم ذلك عن طريق تقليل الأوزان المرتبطة بالمواد غير ذات الصلاحية وزيادة الأوزان للمواد الصالحة.

وعلى الرغم من أن نظام سمارت هو نظام تجريبى، ولم يتم تشغيله على نطاق واسع، إلا أن هناك بعض النظم مثل نظام Cite system الذى وضعه (Dosz Kocs, 1983 ... cited by Lancaster, 1991, p. 244) والذى يحتوى على التغذية المرتدة ذات العلاقة أو الصلاحية والاختصار cite يدل في معناه على Computerized Information Transfer in English وقد استخدم مع نظام Medline فـــى الطب وCatline فى الكيمياء ولعل أهم مميزات نظام Doszkocs هو أنه لا يحتاج إلى حسابات مسبقة لتجمعات المصطلحات.

(٢) بعض البحوث الحاضرة والمستقبلية:

(أ) مصطلحات التكشيف الآلى واستخدامها فى إعداد المستخلصات:

قام الباحث جوناثان كوهين (Cohen, J.D., 1995) باكتشاف طريقة لاستخراج مصطلحات الكشاف من النص، وهذا المدخل لا يستخدم الكلمات المحظورة (قائمة

الاستبعاد no stop list) أو ستيمر Stemmer أو أى مكون لغوى أو موضوعى محدد، وبالتالى يسمح النظام بالتشغيل بأى لغة وفى أى موضوع مع تعديلات طفيفة وتستخدم هذه الطريقة ما يسمى بعدّ إن جرام n-gram Counts وبالتالى تحقق وظيفة مشابهة للستيمر Stemmer ولكنها أكثر عمومية منه، ومصطلحات الكشاف التى يتم توليدها والتى يسميها المؤلف الأضواء Highlights مناسبة لتحديد الموضوع للدراسة والاختيار والأضواء تخدم كمستخلصات خام، كما يصف لنا الباحث امتداداً آخر لاختيار مصطلحات الكشاف التى تمثل مجموعة فرعية من الوثائق حيث تميزها عن الجسد الرئيسى للوثائق، ويقدم لنا الباحث بعض النتائج التجريبية مظهراً إمكانية تشغيل النظام بالإنجليزية والأسبانية والألمانية والجورجية والروسية واليابانية، والباحثة تتمنى أن تتكون فرق بحثية عربية تضم كلا من اختصاصى المعلومات واختصاصى الحاسبات والإحصاء واللغويات لتطبيق مثل هذه التجارب على النصوص العربية ثم تعميمها بعد ذلك.

(ب) بحوث الاستخلاص الآلى:

يذهب الباحث جونسون (Johnson, F., 1995) إلى أن الأمل فى توليد المستخلصات آليا قد جذب العديد من الباحثين ولكن تحقيق تخطى الجهد البشرى فى الاستخلاص مازال أمراً لم يتحقق ويقوم الباحث بمسح مختلف الأساليب لتأكيد أسباب الحقيقة السابقة، وبيان المعوقات التى تحول دون الوصول إلى الهدف، ويقترح الباحث فى هذا الصدد أن التقدم فى بحوث الاستخلاص الآلى يمكن أن يأتى عبر تكامل بعض الأساليب مع النظم المحسبة لاسترجاع المعلومات، وسيسمح ذلك للباحثين بأن يحولوا اهتمام الباحثين من تكرار المزايا التقليدية للمستخلصات إلى تأكيد مزايا التمثيل المحسب للمعلومات فى قواعد البيانات النصية الكبيرة.

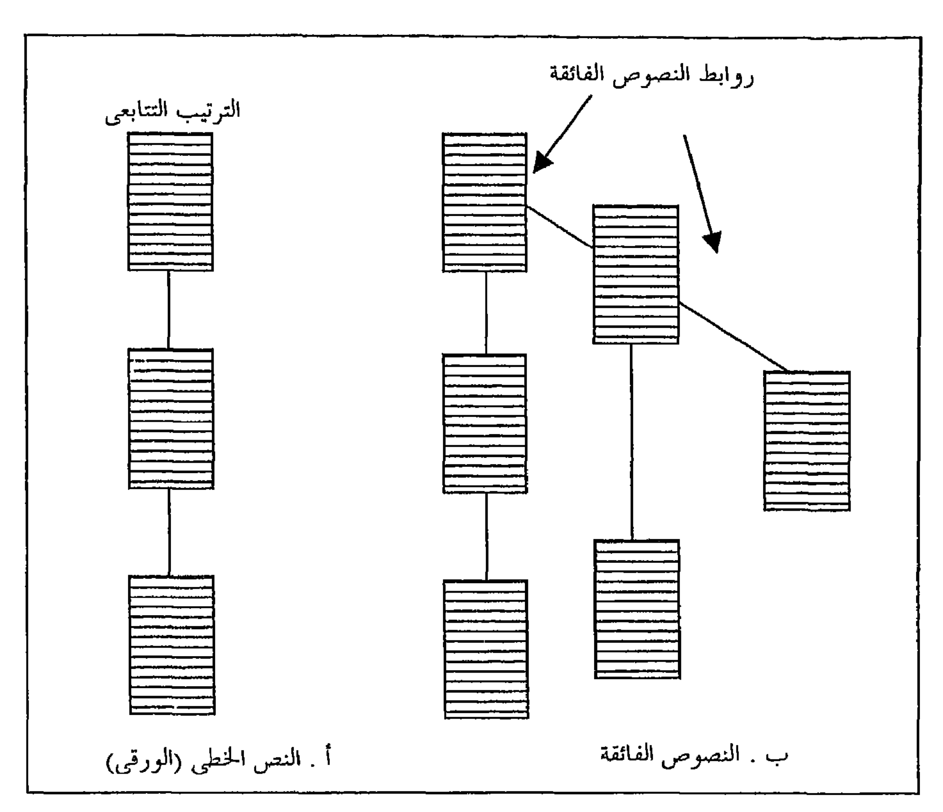
(ج) استيعاب النص أساس للاستخلاص الآلى:

تهتم ورقة بحث الجيهانى (El-Gihani, A., 1995) بتطوير ووضع طريقة لاستيعاب وفهم النصوص العلمية وذلك لتسهيل التوليد الآلى للمستخلصات، ويركز الباحث على جانبين هما الدلالات Semantics والسياق Context، ويعتمد فى هذا العمل على نموذج معرفى Cognitive من ثلاث مراحل لفهم اللغة، حيث يذهب

الباحث إلى أن الفهم ينتج من تمثيل النص رمزياً Symbolically فى الذاكرة طويلة المدى، وقد وضع الباحث طريقة لالتقاط دلالات semantics الجمل، وبالتالى إمكانية عمل مقارنات بين الجمل، وأخيراً فتصف ورقة البحث طريقة لالتقاط السياق الخاص بالنص باستخدام ما يسميه "كيانات الذاكرة" memory entities وذلك لبناء نموذج الذاكرة.

(د) الهيبرتكست وتلخيص وبناء النص الآلى:

تكمن قوة الهيبيرتكست فى الروابط Links التى يقيمها بين أجزاء من النص فى وثيقة واحدة (intra links) أو بين أجزاء من النصوص فى وثائق متعددة (inter links) وبالتالى فالهيبيرتكست أو النص الفائق يقدم لنا مرونة أكبر من النصوص المطبوعة من ناحية تكوين وبناء المعلومات (ناريمان متولى، ١٩٩٧- ص٦ - ٧) كما يذهب حشمت قاسم فى ترجمته لكتاب لانكستر وارنر إلى أن العناصر النصية ترتبط فى النصوص الفائقة بأكثر من طريقة ومن ثم فإن هذا النظام لا يتوافر به نصوص قائمة بذاتها فحسب، وإنما يربط بين هذه النصوص وبعضها البعض بحيث يمكن للمستفيد أن ينفذ إلى ما يريده منها. ثم ذكر الشكل التالى لتوضيح البنية النظرية للعناصر كما تتصل ببعضها فى النظم الورقية ونظم النصوص الفائقة. (لانكستر ووارنر، ١٩٩٧ -ص ٣٩٨ - ٣٩٩).



مقارنة بين الوثائق التقليدية ووثائق النصوص الفائقة

لقد قام الباحث المشهور سالتون وزملاؤه (Salton, G., et al 1997) بدراسة فى هذا المجال، حيث أشار الباحثون إلى أن أساليب استرجاع المعلومات قد استخدمت فى السنوات الأخيرة للتوليد الآلى لروابط النص الفائق الدلالية. وبالتالى فهو وزملاؤه يطبقون أفكار بحوث توليد الروابط الآلية لمواجهة مشكلة هامة أخرى فى معالجة وتجهيز النص وهى تلخيص النص آليا automatic text summarization ذلك لأن أداة تلخيص النص آلياً ستكون ذات منفعة هائلة فى عصر المعلومات المتدفق Information Overload وعن طريق استخدام الجوريزم توليد روابط النص الفائق آلياً، لتوليد الروابط بين الوثائق inter-document links generation أمكن للباحثين توليد روابط داخل الوثائق نفسها intra- document links generation أى بين الفقرات أو المقاطع الخاصة

بالوثيقة، واعتماداً على نموذج الربط داخل الوثيقة، أمكن للباحثين توصيف بنية النص structure of the text ثم قام الباحثون بتطبيق معرفتهم عن بنية النص لعمل تلخيص للنص آلياً عن طريق اقتباس الفقرات passage extraction، وقد قام الباحثون بتقييم مجموعة من خمسين ملخصاً تم توليدها باستخدام هذا الأسلوب ومقارنة هذه الملخصات باقتباسات الفقرات التى تمت بواسطة الإنسان، فتبين نتيجة لذلك أن هذه الملخصات الآلية مناسبة تماما، خاصة على ضوء الحقيقة المعروفة وهى أن الملخصات التى يتم توليدها وإعدادها بواسطة اثنين من الخبراء لنفس المقال تكون غير متشابهة.

ويمكن توضيح فكرة سالتون وزملاؤه بطريقة أخرى، ذلك لأننا يمكن أن نقوم بتحليل البناء الداخلى للنص إلى عناصره المكونة ثم تجميع هذه القطع أو العناصر طبقاً لوظيفتها فى النص، ويمكن استغلال معرفتنا ببناء النص فى إنتاج ملخصات شاملة للنص عن طريق الاقتباس الآلى للفقرات، وعلى ضوء معرفتنا بأن المقتبسات التى يقوم بها يدويا اثنان من الخبراء لنفس المقال تكون مختلفة أى غير متماثلة تماماً، فإن القيام بالتلخيص الآلى للنص بالطريقة السابقة يعتبر أمراً مقبولاً على الرغم من أن هذه الملخصات ليست مثالية، وأخيراً فيؤيد الباحثان هاهن وريمر (Hahn and Reimer, 1984.. cited by Lancaster, F., 1991, p. 246) تطوير النظم الخبيرة المعتمدة على التكثيف Condensation والباحثان يفضلان **تكثيف النص** بدلاً من مصطلح الاستخلاص.

خامساً: الذكاء الاصطناعى والنظم الخبيرة وأهميتها فى عملية الاستخلاص:

أولاً – الذكاء الاصطناعى:

يستخدم مصطلح الذكاء الاصطناعى فى حقل استرجاع المعلومات للدلالة على أى عملية يقوم بها الحاسب الآلى وكان يقوم بها الإنسان مسبقاً، ولكن استخدام كلمة ذكاء هنا غير سليمة فبرامج الحاسبات لا تتعلم من أخطائها ولا تعدل من إجراءاتها بناء على ذلك، وفى الواقع فإن بعض الناس يستخدمون المصطلح للدلالة على مداخل تكرار الكلمات كما استخدمها Luhn منذ أكثر من ثلاثين عاماً، ويذهب كوهلر (Kuhler, 1984) إلى إننا لا نعرف معلومات كافية عن العمليات الفعلية التى تتضمنها عملية الاستخلاص (والتكشيف) حتى يمكن وضع برامج

يمكن محاكاتها بواسطة الحاسوب، وهناك مصطلحات أخرى مرتبطة بالذكاء الاصطناعى مثل نظم الخبرة أو النظم الخبيرة أو النظم المعتمدة على المعرفة، وهذه المصطلحات الأخيرة يمكن تقبلها أكثر من غيرها نظراً لأنها تعنى تطوير قواعد البيانات التى تعكس الخبرة والمعرفة الإنسانية فى مجال معين. ولعل أحد الأمثلة المعبرة عن ذلك هو النظام المصمم للتشخيص الطبى، حيث تعتبر القاعدة المعرفية للأطباء هامة فى ارتباطاتها بالأمراض المختلفة. وفى مجال التكشيف فمن الممكن وضع قاعدة معرفية تعكس قواعد التكشيف وبعض القرارات التى قام بها المكشفون الإنسانيون من قبل (Lancaster, F., 1991, p. 245).

ثانياً- النظم الخبيرة:

هناك اهتمام واسع فى السنوات القليلة الماضية بإمكانيات تطبيق النظم الخبيرة على العمليات المكتبية التقليدية، ويسير هذا الاتجاه متوازيا مع التطورات الحديثة فى حقل الذكاء الاصطناعى، وهناك نشاط يتوقع له النجاح فى هذا الاتجاه ويتلخص فيما يلى:

١- البحث فى قواعد المعلومات على الخط المباشر:

ليس من المستغرب أن يكون هذا واحداً من المجالات الأولى للبحث، ذلك لأن الباحث الجيد على الخط المباشر يمكن اعتباره خبيراً متخصصا، ويتركز النشاط هنا أساساً على نظم البوابات gateways حيث تسمح هذه النظم للمستفيد النهائى بالقيام بالبحث بنفسه. ومما ينبغى الإشارة إليه فى بحث قواعد البيانات على الخط المباشر أن هناك نظما مختلفة تطورت بطريقة مستقلة، على الرغم من وجود جوانب تشابه وجوانب اختلاف فيما بينها بالنسبة لبناء الملفات ولغات الأوامر وبروتوكولات الوصول، ولعل الخبير الوسيط هو وحده القادر على السيطرة على هذه المشكلات فى النظم المختلفة، وبالتالى زاد الاهتمام بالنظم ذات الحدود التفاعلية المشتركة interface systems وهذه بدورها ترتبط بمفاهيم النظم الخبيرة، وهذه وتلك هى محور النشاط الحالى والمستقبلى.

٢- التكشيف:

هذا مجال معقد للغاية ويتضمن معظم الجوانب الفكرية الأساسية لعلم المكتبات والمعلومات، ولعل أفضل النماذج الموجودة هو ذلك الذى يتم حالياً فى المكتبة

الوطنية الطبية والذى يطلق عليه اسم نظام ميدإندكس The MedIndex system ويستخدم هذا النظام نسخة تجريبية من لغة برمجة ليزب LISP حيث يتم التقاط المصطلحات الكشفية للوثائق بواسطة المكشفين الأطباء وذلك لبناء الأطر المعتمدة على المعرفة knowledge based frames وهذا النظام هو نظام خبير حقيقى ولعل مدخله العام أكثر النظم منطقية للمتابعة المستقبلية ليس فى مجال التكشيف وحده بل فى مجال الاستخلاص كذلك، ولعل النظم المعتمدة على المعرفة السابق الإشارة إليها يمكن تقبلها أكثر من غيرها نظراً لأنها تعنى تطوير قواعد البيانات التى تعكس الخبرة والمعرفة الإنسانية فى مجال معين (Lancaster, 1991, p. 245).

٣- الخدمات المرجعية العامة:

هناك مشروعات بحثية قليلة فى هذا المجال، ومعظم هذه النظم ذات طابع معيارى If-Then وهذه تقود المستفيد من خلال سلسلة متتابعة من الاستبعادات elimination حتى العثور على الأداة المرجعية المناسبة، ومن بين الأمثلة الهامة نموذج ANSER MAN وهو نظام خبرة تجريبى يعتمد على الحاسب الشخصى ويستخدم فى المكتبة الزراعية الوطنية، ويسمح هذا النظام للمستفيدين من المكتبة القيام بعملهم المرجعى بأنفسهم عن طريق الحركة من خلال السؤال/ الجواب حتى العثور على الكتاب المناسب، وإحدى المشكلات الواضحة فى معظم هذه النظم هى أنها تتميز بالتخصص الموضوعى العالى فضلاً عن أنها مبنية حول أدوات مرجعية فعلية.

٤- الفهرسة والتصنيف:

هناك نشاط يتم فى الوقت الحاضر لتطبيق النظم الخبيرة على الفهرسة والتصنيف ويبدو ذلك طبيعياً، نظراً لأن النظم الخبيرة والفهرسة يعملان معاً على أساس المعرفة التى يتم التعبير عنها كمجموعة من القواعد، وقد استخدم كل من دافيد وجيمس (David & James, 1984) فى جامعة اكستر ببريطانيا لغة البرمجة برولوج prolog فى محاولة لبناء استخدام القواعد الأنجلو أمريكية AACRI فى نظام خبرة.

٥- مجالات أخرى:

هناك مجالات أخرى عديدة مثل استرجاع المعلومات باللغة الطبيعية وتمثيل المعلومات بالرسم graphic وكذلك عمليات الاستخلاص والتى يمكن أن تكون النظم الخبيرة مفيدة فى مجالها للأسباب التالية:

٥-١ نوعية المستخلصات أمر ضرورى:

تظل حاجة المجتمع العلمى للمستخلصات المبنية جيداً والمراجعات أمراً حاسماً، وعلى الرغم من أن البحث على الخط المباشر قد قام بتثوير الخدمة المرجعية، إلا أننا يجب أن نتذكر دائماً أن العثور على قائمة من تسجيلات الحاسب الآلى فى جزء من الثانية (Nanoseconds) ليس الأمر الحاسم، ولكن الأهم من هذا هو المعلومات الموجودة فى تسجيلات الحاسب، ولعل نظم الخبرة المعتمدة على الحاسبات يمكن أن تكون مفيدة فى إنتاج المستخلصات الفعالة.

٥-٢ تقدم النظم الخبيرة بعض الجوانب المتميزة:

وتتمثل هذه المميزات فى جوانب عديدة منها الجوانب المالية والإنتاجية والتواجد، والمقصود بالجانب المالى أنه عندما يكون النظام الخبير مطبقا على نطاق واسع operational بالحاسب الآلى فسيكون ذلك بتكاليف أقل كثيراً من المطلوب للخبير الذى يتقاضى مرتباً كبيراً، والمقصود بالإنتاجية هنا هو إمكانية اعتبار النظام الخبير كخبير مقيم وبالتالى تزيد الإنتاجية، أما المقصود بالتواجد فهو إمكانية تواجد الخبير فى أماكن جغرافية عديدة فى نفس الوقت.

٥-٣ المداخل الجديدة مطلوبة:

لقد أعلنت المكتبة الوطنية الطبية الأمريكية (وغيرها) الحاجة إلى تحسين خدمات الاستخلاص وتطوير طرق جديدة، ويقترح استخدام النظم الخبيرة كواحدة من الطرق الجديدة.

وليس هناك سر غامض بالنسبة لمفهوم النظم الخبيرة، فهى تعنى ببساطة تصميم وبرمجة الحاسب الآلى للقيام بالأعمال التى يقوم بها الخبير باستخدام الذكاء والخبرة.

وتحتاج مثل هذه النظم إلى مكونات ثلاثة رئيسية أولها أن تكون لديها قاعدة معرفية Knowledge base تلتقط الخبرة كمعرفة إجرائية وثانيهما أن تكون لهذه النظم آلية الاستدلال والاستنتاج تحكم العملية عن طريق اتخاذ القرارات الخاصة بكيفية استخدام معرفة النظام أى أنها تتحكم فى الخطوات المطلوبة لحل المشكلة الجارية وثالثها أن لنظام الخبرة بيانات كمدخلات بالنسبة للمشكلة الجارية المطلوب حلها، وهناك عدد من البرامج المصممة لإنشاء نظم الخبرة والتى تتراوح فى مدى تعقدها.

ويجب أن يكون واضحاً أن النظام الخبير ليس هو الاستخلاص الآلى فى المعنى التقليدى لهذا المصطلح، على الرغم من إمكانية ارتباط الاثنين، فالاستخلاص الآلى يتضمن توليد المستخلصات بالحاسب الآلى مباشرة من نص الوثائق باستخدام الجوريزم مصمم مسبقاً، كما أن معظم نظم الاستخلاص الآلى تعتمد على نماذج إحصائية لكلمات فى الوثائق وهى فى حقيقة الأمر اقتباسات آلية مباشرة من الجمل فى النص، أما الهدف من نظام الخبرة الخاص بالاستخلاص هو التقاط الخبرة ثم الاستعانة بالتجهيز الإنسانى - أياً كانت مستوى تلك الخبرة - خلال عملية كتابة وإعداد المستخلص، أى أن الهدف هنا هو إنشاء التماس بين الإنسان والآلة human machine interface حيث يقسم الجهد الفكرى والآلى فى كتابة المستخلص بين الوسيط الإنسانى ونظام الخبرة المعتمد على الحاسب (Cleveland, 1990, p. 239).

سادساً: بعض النتائج والتوصيات:

(١) بحوث التكشيف والاستخلاص (أو التلخيص أو التكشيف) الآلى تعتمد اليوم على فرق البحث فالباحث سالتون .Salton, G مثلاً الذى كتب فى هذا المجال منذ الستينيات بمفرده أو مع آخرين نراه يكتب مقالاته ضمن فريق بحث فى التسعينيات (١٩٩٦ / ١٩٩٧) وقد أوردت الباحثة فى هذه الدراسة بعض إسهاماته القديمة والحديثة ذات الارتباط والصلة بالجوانب التى عالجتها، أى أن فرق البحث أصبحت ذات أهمية بالغة فى هذا المجال.

(٢) بحوث التكشيف والاستخلاص (أو التلخيص أو التكشيف) الآلى تحتاج إلى تخصصات متعددة فإلى جانب اختصاص المعلومات هناك اختصاص

الحاسبات والاتصالات والإحصاء واللغويات المحسبة وعلم النفس والمتخصصين فى التحليل اللغوى التركيبى Syntactic والدلالى Semantic والمعجمى Lexical ... إلخ.

(٣) على الرغم من الإسهامات المتعددة الأصلية للباحثين خلال الأربعين سنة الماضية فمازالت هذه الجهود داخل المختبرات وفى طور البحوث التجريبية ولم تظهر إلى النطاق التشغيلى operational فى نظم حقيقية بعد، ولعل ذلك يعود بالدرجة الأولى لعدم توفر البرامج التى تسمح للحاسب بفهم اللغة الطبيعية كما يفهم الإنسان، خاصة والحاسب الآلى لا يقوم بالأحكام القيمية التى يقوم بها الإنسان.

(٤) هناك تجارب تمت على اللغة العربية فى نظم التكشيف الآلى لاسترجاع المعلومات وهذه التجارب تبشر بإمكانية تطبيقها على نطاق واسع بالمستقبل ومن الواجب على مراكز البحوث والجامعات القيام بمزيد من تلك التجارب آخذين فى الاعتبار الخصائص المتميزة للغة العربية والبحوث التجريبية العديدة الواردة فى هذه الدراسة.

(٥) هناك حاجة ماسة لتكوين فرق بحث فى التخصصات المختلفة المشار إليها مسبقاً لدراسة الإنتاج الفكرى العربى وبالذات فى التطبيقات الآلية للتكشيف والاستخلاص، والإفادة فى ذلك من كثير من الدراسات الأجنبية التى أوردتها الباحثة وتتصل بلغات عديدة إلى جانب الإنجليزية (انظر على سبيل المثال لا الحصر دراسة كوهين (Cohen, J., D., 1995).

(٦) لابد أن تتضمن دراسات المعلومات والمكتبات بالأقسام الأكاديمية العربية دراسات متعددة الارتباطات interdisciplinary لا سيما على مستوى الدراسات العليا وذلك لإعداد أجيال المستقبل القادرين على فهم ما يدور من تطورات متسارعة فى المجال فضلاً عن إمكانية تطويع تلك الدراسات للغة العربية والإنتاج الفكرى العربى.

الفصـــل التاسع

تقييم المستخلصات والاستخلاص
دراسة في التطور والمعايير والجودة
خلال نصف قرن

مقدمة:

لقد كان السبب الأساسى وراء بداية كتابة المستخلصات هو تمثيل الوثيقة الأصلية بشكل مختصر لقراءته بواسطة الإنسان، ومنذ بداية الخمسينات من القرن العشرين كانت هناك دراسات بجامعة كيس وسترن ريزرف بأمريكا لإعداد مستخلص يقرا بواسطة الآلة (Perry, J. and Kent, A, 1951) سمى بالمستخلص التلغرافى Telegraphic Abstract ، وكان تقييمه بناء على مقارنته بالمستخلص المكتوب باللغة الطبيعية فى عملية الاسترجاع، ودخلت مصطلحات العامل الدلالى Semantic Factor والمدخل التركيبىSyntactic اللغوى فى التقييم .. وتتابعت الإجتهادات التقييمية خلال النصف الأخير من القرن العشرين، ومعظم هذه الإجتهادات كانت نظرية، وكانت النظم التطبيقية قليلة، واعدت المعايير خصوصا تلك التى اعدها معهد المعايير الوطنى الأمريكى (ANSI, 1979) ومعايير المنظمة الوطنية لمعايير المعلومات (NISO, 1997) وغيرها من تعليمات مركز معلومات البحوث التربوية (ERIC)، ومع زيادة التجهيز والمعالجة الإلكترونية للمعلومات وزيادة قوة التحسيب، أصبح بحث النص الإلكترونى هو القاعدة وليس الاستثناء، وإذا كانت هذه الدراسة ستتناول تاريخ تطور تقييم المستخلصات ومستويات جودتها وتطبيقاتها وصفاتها المرتبطة بمنظورات الاستخلاص فسنتناول بعض الاختبارات الكمية الموضوعية للتعرف على مدى جودة المستخلصات فى ثلاث قواعد بيانات، فضلاً عما ظهر فى الإنتاج الفكرى المعلوماتى الحديث عن اكتشاف المعرفة Knowledge discovery فى قواعد البيانات والطرق العاصرة التى تتجاوز مجرد بحث النصوص والاستخلاص الآلى إلى روابط النص وتوليد النصوص والاستخدام الذكى لمعالجة النصوص (Lancaster, F., 1999).

أولاً: تعريف المستخلص الجيد وأركانه (ناريمان متولى، ١٩٩٨، ١٠٣ مع التحديث):

حتى يمكننا الحكم على نوعية المستخلصات وجودتها فمن الضرورى، أن نتعرف على ما يجعل هذه المستخلصات جيدة، وهناك خطورة فى إصدار الأحكام بأن هذا الشيء جيد أو غير جيد، ولكن بالنسبة للمستخلصات هناك بعض الاتفاق العام بين الخبراء على معايير الجودة أو النوعية. فهناك معايير المستخلصات التى أصدرها معهد المعايير الوطنى الأمريكى (ANSI) عام ١٩٧٩ والتى تؤكد على نوعية المستخلصات الجيدة حيث يعرفها المعهد المذكور بأنها (تمثيل مختصر ودقيق لمحتويات الوثيقة). أما كوليزون (Collison R., 1971, P. 3) فيصف المستخلص بأنه:

(التمثيل الموجز المحكم Terse لجميع النقاط الموجودة فى المعلومات الوثائقية الأولية الأصلية على أن تكون بلغة المؤلف وبنفس الترتيب التى جاءت به فى الوثيقة الأصلية، والمستخلص فوق ذلك لابد أن يمثل كياناً مستقلاً مقروءاً وكاملاً وبالتالى فهو عمل علمى منفصل يمكن تكشيفه والإفادة منه لخدمة الباحثين).

١-١ ما الذى يشمله المستخلص؟

لعل الإجراءات التى تتخذ لكتابة المستخلصات يمكن أن يكون لها علاقة بالجودة أو النوعية الجيدة للمستخلص فى النهاية، وطبقاً لما يذهب إليه كليفيلند (Cleveland, 1990) فإن القائم بالاستخلاص يقرأ أو يفحص الوثائق للبحث عن خمسة معينات عن المحتوى وهى:

(أ) الأهداف.

(ب) المنهج.

(ج) النتائج Findings & Results.

(د) الخاتمة Conclusion.

(هـ) أى معلومات إضافية حول الموضوع أو تدعمه.

وواضح أنه باتباع هذه الإجراءات فإن المقالة البحثية المحكمة البناء، تكون أسهل فى استخلاصها من مقال الرأى أو المقالات المنشورة فى المجلات ذات الاهتمام العام. وهناك تعليمات وضعتها إريك (ERIC) حتى يلتزم بها القائمون بالاستخلاص عند قراءتهم للوثيقة بغرض الاستخلاص (ERIC, P. 11) وهذه التعليمات تشمل ما يلى:

(١) المادة الموضوعية ونطاقها وأهدافها.
(٢) نوع المطبوع/ الوثيقة.
(٣) وجهة نظر المؤلف.
(٤) الجمهور المستهدف.
(٥) علاقة هذا العمل بالأعمال الأخرى.
(٦) الاستخدام المقصود.
(٧) بعض الخصائص المميزة كوجود قائمة مصطلحات أو خرائط... الخ.
(٨) النتائج أو ما توصل إليه الباحث.

هذا ولا تتضمن المستخلصات عادة كما سبق أن ذكرنا الخلفية التاريخية، المقدمة، الملخصات، تكرار المعلومات، المعلومات القديمة، تفاصيل إجراءات بحثية معيارية، معلومات من المتوقع أن يعرفها القارئ، أفكار عن البحوث المستقبلية، بيانات خام، كل النتائج، التحيزات الشخصية أو الآراء الخاصة بالقائم بالاستخلاص.

ولعل أحد الأوجه الرئيسية لجودة أو نوعية المستخلص هو وجود النقاط الرئيسية للوثيقة (Lancaster, 1991, P. 105) وهذا الوجه يتأثر بطريقة مباشرة بالتطبيق الدقيق للقواعد المرشدة السابق الإشارة إليها، أو أى إجراءات مشابهة لتوضيح المعلومات اللازمة للمستخلص.

ويقترح لانكستر فى الطبعة الثانية من كتاب التكشف والاستخلاص (Lanccster, F., 1998) اثنين من المعايير الشاملة للحكم على جودة المستخلص وهى تتلخص فى الاجابة على السؤالين التاليين:

أ- هل تم تغطية النقاط الرئيسية فى المقال.

ب- هل تم تمثيل هذه النقاط بدقة ووضوح.. وعلى الرغم من أن معايير المنظمة الوطنية لمعايير المعلومات (National Information Standards Organization, 1997) تقدم لنا توجيهات عن الأسلوب إلا أنها لا تزودنا بمعايير لتقييم النوعية والجودة.

١-٢ تقييم المستخلصات: بين الذاتية والموضوعية:

تقييم المستخلصات فى معظم الأحيان هو تقييم ذاتى (جيد/ سيئ) وقد تستخدم بعض الأدوات الكاشفة للأخطاء مثل الإستشهادات المرجعية الصحيحة، الوصف الحقائقى أو حذف بعض النقاط الهامة، وقد ننقد الأسلوب والنحو والفقرات المتكررة. وبعض التقييم يأتى من مدى انتظام تطبيق سياسة وقواعد الاستخلاص، وسرعة النشر وإتاحة المستخلصات والتكاليف والثقة والسمعة فيمن يقومون بإعدادها ويدخل كذلك عامل التكلفة فى الحسبان، فالاستخدام يتأثر بعدم إمكانية الإتاحة الاقتصادية، وكثيراً ما يشكو الأمناء من ارتفاع أثمان خدمات التكشيف والاستخلاص المستمر وضرورة إلغاء بعضها من اشتراكات المكتبة لعجز الميزانية. والمحك النهائى لجودة المستخلصات هو مدى استجابة تلك المستخلصات لتلبية الاحتياجات المعلوماتية للمستفدين (Cleveland, 1990, P. 197.) .

١-٣ نوعية الاستخلاص وانتظامه : Consistency

من الملاحظ أنه ليس هناك مستخلصين متشابهين لوثيقة واحدة ويظهر ذلك إذا اعد المستخلص شخصين مختلفين أو قام بهما شخص واحد ولكن فى فترات متباعدة، وعلى ذلك فإن نوعية وانتظام المستخلصات تظهر فى الرد على السؤالين التاليين:

(١) هل تضمن المستخلص النقاط الرئيسية فى الوثيقة.

(٢) هل تم وصف هذه النقاط بدقة، وبتحديد، أى أن نوعية الاستخلاص يمكن الحكم عليها طبقا لمعايير مشابهة لتلك المستخدمة فى تقييم التكشيف، ذلك لأن الخطوة الأولى فى الاستخلاص كما هو الحال فى التكشيف هو التحليل المفهومى. أما الخطوة الثانية، فهى ترجمة هذه المفاهيم إلى نص مقروء.

أما عملية الانتظام فى إعداد المستخلصات فيمكن الحكم عليها على المستوى المفهومى إذا ما تم اتفاق المستخلصون على النقاط التى يجب أن يشملها المستخلص.

ومن الواضح أن تقييم المستخلصات بواسطة المحررين وغيرهم الذين يعملون داخل مراكز المعلومات أو بيوت الناشرين، هذا التقييم يتم بالطريقة الذاتية المنطقية، ويبدو أنه مع زيادة استخدام بحوث النصوص الأساسية بدلاً من التكشيف الإنسانى، فإن المدخل الإسترجاعى للتقييم يصبح أكثر أهمية.

١-٤ قضايا الملاءمة Compatibility Issues

إذا كان السبب الوحيد فى كتابة المستخلصات منذ حوالى ثلاثين عاماً هو إعداد تمثيل للوثيقة يتم قراءته بواسطة الباحثين أنفسهم، فإن المستخلصات فى الوقت الحاضر تكتب لخدمة غرض إضافى، وهو تقديم وإعداد التمثيل للوثيقة حتى يمكن البحث عنها بواسطة الحاسب الآلى، ولسوء الحظ فلا يتطابق هذان الهدفان بطريقة متلائمة متماثلة تماماً، فقد يكون التكرارRedundancy مرغوباً لخدمة أغراض الاسترجاع، أى أن الموضوع نفسه يمكن تمثيله بأكثر من طريقة واحدة، وعلى سبيل المثال، فإن وجود المترادفات فى بعض المستخلصات سيزيد من احتمالات استرجاع المادة نفسها (Data Wings = Triangular Wings وبالتالى فقد يستخدم أحد الباحثين الصفة الأولى، بينما الأخر يستخدم الصفة الثانية) كما ينبغى أن نلاحظ أن الإنسان القارئ من جهة أخرى يفضل ويرغب فى الانتظام Consistency أكثر من رغبته فى التكرارRedundancy. وفى واقع الأمر فقد يختلط الأمر كثيراً على المستفيد إذا ما تم وصف الأفكار بطرق مختلفة داخل المستخلص الواحد. وكلما كان المستخلص أطول، كان أفضل لخدمة أغراض الاسترجاع، ذلك لأن هذا المستخلص سيزودنا بنقاط أكثر للوصول، وكلما زادت نقاط الوصول كلما ارتفع الاستدعاء Recall فى عملية الاسترجاع وفى نفس الوقت فيجب الاعتراف بأن الدقة Precision ستقل، وذلك لأنه كلما كان المستخلص أطول كلما توفرت جوانب ذات أهمية أقل فى الوثيقة، أى ظهرت هذه الجوانب كنقاط وصول، ومعنى ذلك زيادة الارتباطات الكاذبة، أى أن المستخلص المختصر مرغوب للباحث بصفة عامة، كما أن هذا المستخلص المختصر مرغوب للمشتركين فى الأدوات المطبوعة. ذلك لأن المستخلصات الأطول تؤدى إلى

مطبوعات أكثر تكلفة. كما يمكن أن نلاحظ أيضاً أن هناك بعض الكلمات التى ينبغى تجنبها لأغراض الاسترجاع، وعلى سبيل المثال فاستخدام كلمة Aids بمعنى يساعد، ستخلق مشاكل فى العديد من قواعد البيانات، لأنها ستؤدى إلى استرجاع مواد عن مرض الإيدزAids، بينما كلمة يقودto lead بالإنجليزية ستؤدى إلى استرجاع بعض المواد عند البحث تتصل بالمعدن lead وهو الرصاص أى أنه لابد للقائمين بالاستخلاص من تجنب استخدام بعض الكلمات التى ستؤدى إلى مشكلات من هذا النوع إذا أرادوا استرجاعاً أفضل. وحتى علامات الترقيم والتركيب اللغوى والذى له دلالته بالنسبة للقارئ، قد يسبب مشكلات للحاسب الآلى، فإذا كانت هناك جملة تنتهى بكلمة حامض Acid، وتليها مباشرة فى الجملة التالية كلمة ترسيب Precipitation فقد يؤدى ذلك فى بعض النظم إلى استرجاع مواد عن Acid Precipitation على الرغم من أن الوثيقة والمستخلص لا يتصلان نهائياً بهذا الموضوع.

أما بالنسبة للمستخلصات المصغرة Mini Abstracts والتى أنشأها الباحث لونين Lunin فهى مصممة بالدرجة الأولى لتيسير البحث بالحاسب الآلى، وعلى الرغم من أن مثل هذه المستخلصات يمكن تفسيرها بالمستفيدين الأذكياء، فإنها بالقطع أكثر صعوبة فى القراءة والفهم. لقد كانت هذه النقاط أو الجوانب السابقة إشارة إلى أن المستخلصات المثالية أو المناسبة للإنسان القارئ ليست هى بالضرورة المناسبة للبحث بواسطة الحاسب الآلى، وإن كانت هذه المستخلصات ستخدم خلال المستقبل القريب لخدمة الغرضين معاً، وحتى لو قلت أهمية الخدمات المطبوعة، فستظل الحاجة قائمة للمستخلصات كمخرجات وسيطة فى البحوث المعتمدة على الحاسب الآلى.

ولعل أحد النتائج لهذه الدراسات السابقة يتعلق بناشرى الخدمات الثانوية، بحيث يقوم هؤلاء بمراجعة تعليماتهم للمستخلصين، حتى يقوموا بإعداد المستخلصات التى تكون بديلاً مناسباً للوثيقة ولخدمة كل من البحث المحسب والقراءة.

ولقد قامت فيدل (Fidel, 1986) بخدمة كبيرة بتحليلها لتعليمات الاستخلاص التى قام بإعدادها عدد (٣٦) منتج لقواعد البيانات، كما قام لانكستر Lancaster فى التسعينات بتلخيص هذه التعليمات بحيث تصبح صالحة للصفات الإسترجاعية للمستخلصات.

ولعل الباحثة تختم هذا الجزء عن التعريف وأركانه بأكثر القوائم شمولا فى تقييم المستخلصات والتى أعدها الباحثان (Barko & Bernier) والتى تتضمن المعايير التالية:

(١) الترتيب المنهجى لنوعية المستخلصات (ويتم ذلك بواسطة محكمين من العلماء).

(٢) مدى تطبيق معايير ASNI أو غيرها من المعايير (ويعتبر هذا البند مكون أساسى فى مدخل التقييم الذى يوحى به الباحث Mathis).

(٣) مدى شمول المستخلص للمعلومات ذات الدلالة ومدى استبعاده للمعلومات غير الهامة.

(٤) عدم وجود الأخطاء.

(٥) انتظام الأسلوب والخاصية القرائية Read ability.

(٦) صلاحية التنبؤ Relevance Predicatability.

(٧) إمكانية اعتبار المستخلص كبديل Surrogate للوثيقة الأصلية (مستخلص إعلامى).

(٨) درجة كفاية المستخلص كمصدر للمصطلحات الكشفية.

ومن الواضح أن هذه القائمة تمثل مستويات مختلفة من المعايير، من رقم ٣ – ٥ يفترض أنها تنسحب على أى تقييم أو ترتيب كونى، أما تقييم مدى الاستعانة بالمستخلص كبديل للوثيقة الأصلية (المعيار رقم ٧) فهذا المدخل يتصل بمقارنة قدرة الأفراد القائمين بعمل المستخلصات على الإجابة على الأسئلة المعتمدة على كل من المستخلصات والنصوص الكاملة.

ثانياً: تاريخ تطور تقييم المستخلصات خلال نصف قرن:

يتكامل التكشيف مع الاستخلاص، وإذا اعتبرنا التكشيف الخطوة الأولى لإعداد المستخلصات، أو إذا اعتبرنا أن كلا من التكشيف والاستخلاص يتضمن إعداد تمثيل للمادة الموضوعية فى الوثائق، أو إذا اعتبرنا أن الغرض الأساسى للمستخلص هو بيان ما تتحدث عنه الوثيقة Aboutness، فإن عددا مصطلحات

الكشاف، يمكن أن يخدم نفس الغرض، بمعنى أن قائمة المصطلحات فى مجال معين، يمكن أن يعتبر مستخلصا صغيرا Mini- Abstract، ولقد جاء تقييم التكشيف فى عدد من الدراسات العربية (احمد بدر ١٩٩٨) (حشمت قاسم ٢٠٠٠ ص ١٤٣- ١٦٣)، أما بالنسبة لتقييم المستخلصات فكانت المعالجة قليلة بل نادرة فى الدراسات العربية والأجنبية، ومن هنا كان حرص الباحثة على محاولة ملء هذا الفراغ وإن كان تقييم الكشافات ليس بعيدا تماما عن تقييم المستخلصات ولعلنا نرجع هذا التاريخ إلى مؤتمر المعلومات العلمية الذى عقدته الجمعية الملكية البريطانية عام ١٩٤٨ حيث قدمت فيه أوراق بحثية عن نطاق ونوعية المستخلصات وقد أوصى المؤتمر بأن يقوم المؤلف بإعداد المستخلص بنفسه فى مجالات العلوم والتقنية، للارتفاع بنوعية المستخلصات وجودتها ومستواها، كما جاءت أوراق بحوث أيضا عن فاعلية خدمات التكشيف والاستخلاص ضمن وقائع المؤتمر الدولى للمعلومات العلمية الذى عقد بواشنطن عام ١٩٥٨، وقد حفلت فترة الخمسينات بدراسات فى جامعة كيس وسترن ريزرف بأمريكا عن المستخلصات التلغرافية (وهى مستخلصات آلية)ومقارنتها بالمستخلصات الطبيعية وادخل فى أدب المكتبات منذ ذلك الوقت طريقة التحليل المعروفة باسم العامل الدلالى: Semantic Factor وظهرت أهمية تقييم المستخلصات من الناحيتين اللغوية والدلالية (Perry, J. Semantic and Syntactic approaches W., 1951) وفى عام ١٩٥٩ قام الباحث ادمندسون وزملاؤه (Edmundson et al, 1959) باقتراح عدة قواعد Criteria لتقييم المستخلصات من بينها: المقارنة مع مستخلصات مثالية Ideal، مدى استرجاعية Retreievability الوثيقة بالمستخلص، ومدى إمكانية استخدام المستخلص للإجابة على أسئلة اختبار فضلا عن استخدام الاحكام الذاتية، وقد اقتراح الباحث باين وزملاؤه (Pinto, M., 1999, 239) اختبارا لجودة المستخلصات عن طريق إجابة المستخلص على عدد من الأسئلة فضلا عن استخدام اختبار الانتظام Consistency والذى يتم فيه مقارنة التشابه بين المستخلصات المختلفة والتى تم اعدادها من نفس الوثيقة بواسطة عدد من المستخلصين.

أما الباحث فينسونهالر (Visonhaler, 1966) فقد أوصى: تنطبق مقياس ترتيبىScale ذو سبعة درجات لتحديد مدى التشابه بين المستخلص والوثيقة

الأصلية، أى أنه اقترح بعض الأساليب السلوكية للحكم على نوعية المستخلصات بناءً على صحة محتوياتها أو صحة تنبؤاتها، وفى دراسات صحة المحتوى Content- validity فإن الموضوعات هى التى تحدد درجة تشابه الوثيقة مع المستخلص الخاص بها، وذلك باستخدام مقياس التشابه ذى النقاط السبع التى اقترحها، وهناك طرق بديلة لعمل اختبارات تحدد مدى تمييز المستخلص بين الوثائق المختلفة التى تبدو فيها المادة الموضوعية متشابهة إلى حد كبير.

وقد قام فينسو نهيلر Vinsonhaler أيضاً باقتراح اختبار لقياس هذا التمييز. وهناك اختبار آخر للصحة التنبؤية، وهو اختبار أكثر تقليدية والذى يدل على مدى التنبؤ الصحيح للمستخلص عن صلاحية الوثائق.

أما العالم ماثيس(Mathis,1972) فقد رأى إمكانية تقييم المستخلصات بناءً على معامل البيانات Data Coefficient، ويتم التعبير عن معامل البيانات هذا بواسطة المعادلة $\frac{C}{L}$ حيث يعتبر الرمز C هو عامل حفظ البيانات، والرمز L هو عامل الطول والمقصود بالحرف C أنه قياس للمدى الذى تكون فيه جميع المفاهيم الخاصة بالوثيقة موجودة فى المستخلص (ومقصد ماثيوس بالمفاهيم هنا عناصر البيانات)، أما الحرف L فهو يدل ببساطة على كلمات المستخلص مقسومة على عددها فى الوثيقة. ومعنى ذلك أن معامل البيانات (DC) يدلنا على قيمة رقمية، وكلما كانت هذه القيمة أعلى كلما كانت أفضل، ذلك لأن المستخلص فى هذه الحالة سيعكس خاصة التركيز والقدرة على الاحتفاظ بجميع العناصر الأساسية للنص بأقل الكلمات، كما يتم تحسين قيمة المستخلص إما بزيادة عدد عناصر البيانات الموجودة أو تقليل عدد الكلمات فى المستخلص، ويرى الباحث ماثيس أن قيمة معامل البيانات التى تقل عن (١) ستدل على نوعية غير مقبولة من المستخلص.

وقد استخدمت العديد من هذه المداخل عبر السنين، وكان أكثرها تفضيلا لدى الدارسين فى المجال هو اختبار قدرة المستخلص للتنبؤ بصلاحية relevance الوثيقة لاحتياجات معلوماتية محددة، وكان من بين هؤلاء كنت (Kent, A, et al,

(1967 الذى قام مع زملائه باختبار وتقييم معالجة المعلومات الالكترونية، وحتى هارتلى (Hartely, J. et al, 1996) الذى قدم لنا دراسة يتم الحكم فيها على مقدرة المستخلصات على الإجابة على الأسئلة المختلفة وفى هذه الحالة فالمقارنة تتم بين مستخلصات مركبة Structured بأخرى غير مركبة Unstructured.

أما سالتون وزملاؤه (Salton, G, et al, 1997) فقد استخدم عددا من مداخل التشابه similarity approaches وهو الذى يعنى مدى تشابه الاقتباس الآلى مع الاقتباس الإنسانى وأخيرا فقد درس ويتلى (Wheatley, A, 1997) ما يسمى بالقرائية readablity لمختلف المستخلصات المعتمدة على مصادر الإنترنت.

وستورد الباحثة فى الجزء التالى من الدراسة توضيحا وتعريفا لمختلف مستويات الجودة وصفاتها المرتبطة بمختلف منظورات المستخلصات والاستخلاص.

ثالثاً : مستويات الجودة والنوعية بالنسبة للمستخلصات :

يستخدم مصطلح النوعية أو الجودة فى حياتنا اليومية كفكرة عامة تعكس المستوى الأعلى، ولكن المصطلح فى استخدامه بالتطبيقات الخاصة بالعلوم والتكنولوجيا والتجارة والتعليم يعنى كيفية القياس والتعبير الكمى عندما تكون الأشياء مادية محسوسة، كما هو الحال فى المنتجات المصنعة ومدى تطابقها مع معايير standards موضوعة، أما فى الأشياء الأقل محسوسية فيصعب تحديد النوعية (مثل نوعية الخدمات المكتبية المعلوماتية فالحكم هنا ذاتى غير موضوعى).

وهناك مستويات عديدة للجودة وضعها بنتو ولانكستر فى الجدول التالى رقم(١)

المنظور	أساس الحكم	الصفات
المستخلص	الفلسفة التأمل	مطلق ساكن
المنظور التنظيمى العملى	معايير ولوائح وقواعد	بعض العمليات يمكن تطبيق اللوائح عليها بإحكام والبعض

		الآخر لا يحقق ذلك
المنتج	المعايير	بعض الأولويات يمكن أن تكون موضوعية ويتم الالتزام بها.
الخدمة	معايير أو قواعد	ذاتية اكثر منها موضوعية ومن النادر إمكانية الالتزام بها.
المستفيد/ الزبون	التكلفة القيمة النظام القيمى الشخصى	ديناميكية نسبية

فمن اللازم إذن الإشارة إلى أن الجودة ليست فكرة واحدة، فهناك مستويات عديدة أو منظورات مختلفة، حيث يوجد فى أحد أطرافها المستخلصات أو الفكرة السامية عن النوعية، وهى فكرة ساكنة، مطلقة، وتوجد فقط فى المنظور الفلسفى الميتافيزيقى.. وفى الطرف الآخر يوجد منظور (المستفيد) وهو منظور شخصى أو ربما فطرى idiosyncratic وهو ديناميكى ونسبى، بمعنى أنه يتضمن دائما مقارنة، واختيار واحد من البدائل المتعددة.. وعادة يتم هذا الاختيار انطلاقا من التكاليف المادية، أو انطلاقا من الزمن الذى تستغرقه الإجابة أو الراحة فى الوصول إلى المعلومات.

هذا ويلاحظ أنه فى العديد من المنتجات التى يجب أن تصنع لتتلاءم مع المعايير، يمكن اعتبار نوعيتها قريبة من صفة (المطلق). أما المنتجات الفكرية كالاشكال العديدة من المطبوعات فهى أقل عرضة للمعايرة الحقيقية (الوعاء هنا هو الورق، التجليد وهذه الأوعية قابلة للمعايرة).

أما العملية فهى غير متجانسة، فبعض العلميات يمكن معايرتها، بل يمكن أن تخضع أحيانا للقواعد المطلقة (كالأمان وغيرها من القضايا الصحية)، ومرة أخرى فالعمليات الفكرية اقل عرضة للمعايرة الحقيقية.

أما بالنسبة *لمنظور الخدمة* فهو يقع بين منظور المنتج ومنظور المستفيد، فالخدمات نادرا ما يمكن الحكم عليها بمعايير مطلقة، وأن كان بعض جوانب الخدمة يمكن التعبير عنه كميا (عدد المقاعد بالنسبة للقراء، عدد الطلاب بالنسبة لكل عضو هيئة تدريس.. فالمعايير هنا نادرا ما يمكن الالتزام بتطبيقها. ومع ذلك ففى خلال السنوات الأخيرة، تم تبنى مبادئ ادارة الجودة الشاملة(TGM) والتى تتضمن التركيز على إرضاء المستفيد وعلى التحسين المستمر.

٣- ١ اعتبارات الجودة فى تطبيقها على الاستخلاص :

يعكس الاستخلاص عملية أكثر طموحا وتعقيدا من مجرد المنظور اللغوى النفسىPscholinguistic ولا ينبغى أن يتم تحليل نص الوثائق ببعض التفصيل ولكن النص (المستخلص) يجب أن يتم إنتاجه بكفاءة. أى أن النص يجب أن يكون متماسكا نحويا ودلاليا Syntactically Semantically بالإضافة إلى ضرورة كونه ملخصا معبراً عن للوثيقة الأصلية.

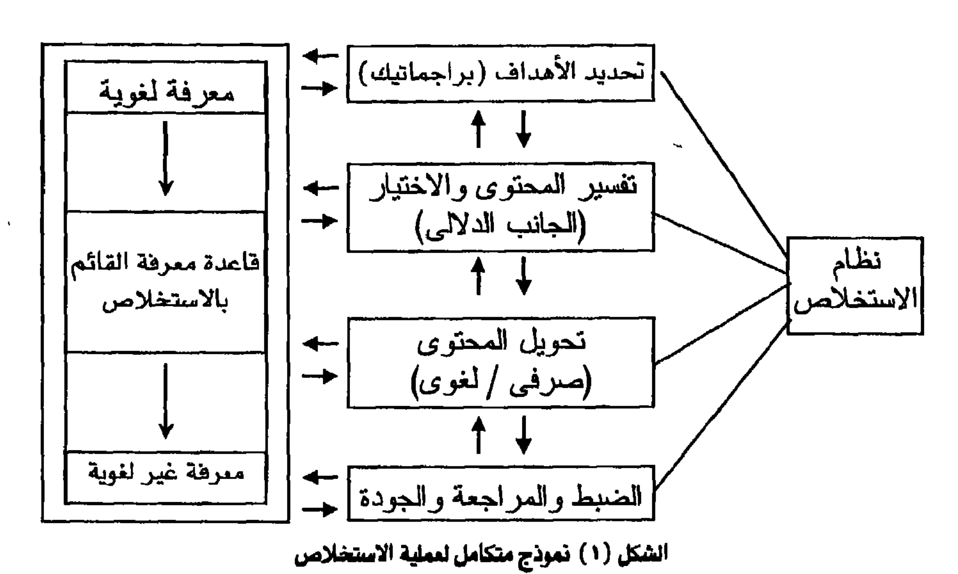
ويعتبر الاستخلاص أصعب العمليات فى بيئة معالجة وتجهيز الوثيقة، ذلك لأن المستخلص فى الوقت الحاضر على الأقل يجب أن يعكس كلا من وصف المحتوى وأداة الاسترجاع. وقد أوضح لنافيدل (Fidel, 1986) أن هذين الجانبين يمكن ألا يكونا متطابقين تماما كما سبقت الإشارة.

ويمكن فى الشكل التالى رقم (١) التعبير عن نموذج لعملية الاستخلاص، ويذهب لانكستر (Lancaster, 1998) إلى أن هناك أربع مستويات للمعالجة والتجهيز حيث يعكس المستوى الأول تحديد الأهداف والغايات بالخدمة أو الدورية التى تنتج المستخلصات، وهذه تنعكس فى القواعد المرشدة التى توجه القائم على الاستخلاص، حيث تتساوى العمليتان التاليتان فى الأهمية وهما:

أ- تفسير المحتوى والاختيار.

ب- تحويل المحتوى أو ما يسمى أحيانا ترجمة هذا المحتوى إلى لغة التكشيف..

فالعملية الأولى تهتم بفهم النص الأصلى وتحديد العناصر التى يجب أن يتضمنها المستخلص، أما العملية الثانية فتهتم ببنية المستخلص أى كيفية تقديم العناصر المختارة فى نص المستخلص.



الشكل (١) نموذج متكامل لعملية الاستخلاص

ونلاحظ فى الشكل السابق أن رأس الموضوع (الضبط والمراجعة) هو الذى يعبر عن العملية المرتبطة مباشرة بالجودة أو النوعية، وهذه العملية لها جوانب عديدة من بينها قيام القائم بالاستخلاص بفرض وجهة نظره عن النوعية قبل تقديم المستخلص لمتابعة عملية المعالجة، وقد يتم مراجعة عمل المستخلص بواسطة المحرر أو بواسطة مستخلص رئيسى قبل النشر، كما قد يستخدم القراء ضوابطهم عن النوعية والجودة والمتعلقة بفهمهم للمستخلص وقيمته بالنسبة لتوقع صلاحية المادة الأصلية لاهتماماتهم الخاصة. ويشير الشكل (١) إلى أن نوعية المستخلص ستتحدد إلى حد كبير– بنوعية قاعدة معرفة المستخلص وهذه المعرفة تشمل كلا من المعرفة اللغوية (أى مقدرته على تفسير لغة النص فى المجال الموضوعى) والمعرفة غير اللغوية أى مدى فهمه للاحتياجات والاهتمامات المعرفية للجمهور المستهدف.

وهناك من الكتاب الذين يركزون على فن تلخيص النص أو على المهارات المطلوبة من المستخلص الجيد(Endres- Niggemeyer, 1995) كما أن تطبيق هذه الجوانب السابقة فى الاسترجاع يزيد من أهمية المستخلصات خصوصا فى بيئة الانترنت (wheatley, 1997).

٣-٢ التعريف بصفات الجودة المرتبطة بمختلف منظورات المستخلصات والاستخلاص:

وضع بنتو ولانكستر (Pinto, M, 1999) جدولا تحليليا لمختلف المداخل ووجهات النظر التى تم اتباعها خلال نصف القرن الأخير والتى تتمثل فيما يلى (الجدول رقم ٢)

منظور العملية	منظور الخدمة
الشمولية Exhaustivity	رضى المستفيد
الدقة Accuracy	فاعلية التكلفة
القرائية Readability	منظور المستفيد
التماسك Cohesion/ Coherence	التكلفة
التكاليف	القيمة
منظور المنتج	منظور العملية/ المنتج
الانتظام Consistency	الكثافة
الاختصار Brevity	التكلفة
التكاليف	

❖ فمنظور العملية:

يتعلق أساسا بالصفات المتصلة بالتمثيل المعرفى، والتشابه هنا يمكن أن يتم بين عمليتى التكشيف والاستخلاص(Lancaster, 1998)، فالشمول يتصل باتساع التغطية، وفى واقع الأمر فهو مقياس لمدى تمثيل النص الأصلى فى المستخلص (خصوصا الأفكار والمفاهيم والخاتمة..).

وفى الأحوال المثالية فإن المستخلص يجب إن يفصل طبقا لاحتياجات جمهور معين، وهذا واضح فى المستخلصات التى تعد لخدمة الباحثين فى شركة متخصصة أو مؤسسة بحثية. أى أن المستخلص هنا يستجيب لاهتمامات مجتمع محدود.. وبالتالى فكلما كان المجتمع الذى تخدمه المستخلصات غير متجانس heterogeneous كلما كان من العسير الوصول إلى اتفاق بالنسبة للعناصر التى يشملها المستخلص.

❖ أما منظور الدقة:

فيشير إلى المدى الذى يمكن أن يمثل فيه المستخلص النص الأصلى بدرجة صحيحة (فقد يخطئ القائم بالاستخلاص بالنسبة لتفسير معنى النص) فهذا خطأ فكرى وقد يكون هناك خطأ إهمال (أى يسجل بيانات خاطئة)، وقد يكون الخطأ ناتجا من عدم إتقان القائم بالاستخلاص للغة المكتوب بها النص.

❖ أما منظور القرائية Readability

فهو يعنى أن المستخلص سيتحدد بمقدرة المستخلص على التعبير عن نفسه بوضوح وتحديد وبدون غموض، مقتديا بالتوجيهات الموضوعة كمرشد للمستخلص Guidelines (فالبعض يرى مثلا أن المستخلصات المركبة structured فى فقرات وكل فقرة لها رأس موضوع، هذه المستخلصات تكون اسهل فى الفهم) وهناك اختبارات لقرائية النص (مثل معادلة فليش Flesch Reading Ease Formula) أو الفهم (مثل معايير كلوز Cloze Criteria)، وإذا ما تم تطبيق هذه المقاييس على المستخلصات، فالقرائية يمكن أن تكون مقياسا موضوعيا ويمكن أن يتم التعبير عنه كميا.

أما التماسك فهو يتصل بالقرائية ولكنه ليس مطابقا لها، أى أنه يمكن أن يتم تقييم المستخلصات شأنها فى ذلك. شأن أى نوع آخر من النصوص على أساس خاصية القرائية Readability وذلك باستخدام معادلة القراءة المعيارية. ومن بين الباحثين فى هذا المجال العالم كنج (King, 1976) الذى استخدم هذه الطريقة لتقييم مستخلصات نمو الأطفال Child Development/Abstracts، وتبين له أن هذه المستخلصات ذات معدل قرائى متدنى بالنسبة للعناصر التى اعتمدت عليها، والتماسك هنا يعكس الوصل بين مختلف أجزاء النص، ويلاحظ أن الاقتباسات Extracts المعدة بواسطة الحاسب (اختيار جمل على أساس احصائى أو حسب معايير مكانها فى النص أو وضعها التركيبى اللغوى) ستكون بعيدة عن صفات القرائيه على الرغم من أن الاقتباس الكامل يمكن أن يكون تمثيلا مرضيا للاتجاهات الأساسية للنص.

❖ أما منظور الإنتاج

فيتصل بالكفاية الفنية للمستخلص، وفكرة الانتظام فى المستخلصات شبيهة بالانتظام فى التكشيف الموضوعى، ذلك لأنها تشير إلى الدرجة التى يقوم فيها اثنان من الأفراد بإنتاج مستخلصات متشابهة (انتظام بينى للمستخلصين)، أو أنها الدرجة التى ينتظم فيها الفرد الواحد فى إنتاج مستخلصات متشابهة فى ظروف مختلفة (انتظام داخلى للمستخلص).

ويجب أن نميز فى موقف التكشيف بين الانتظام فى التحليل المفهومى والانتظام بالنسبة لترجمة التحليل المفهومى إلى مصطلحات (مأخوذة مثلا من المكنز).. ومع ذلك فالانتظام فى الاستخلاص ينطبق فقط على المستوى المفهومى نظرا لأنه من غير المتوقع أن يستخدم الأفراد المختلفين نفس الكلمات تماما أو نفس التركيبات النحوية. وعلى كل حال فمن الواضح أن المستخلص المركب Structured سيكون اكثر انتظاما من غيره.

هذا ويعتبر الاختصار Brevity صفة مرغوبة فى المستخلص الجيد، وهى صفة تخضع للقياس، أما الطول فهو واحد من الصفات القليلة التى تحرص عليها المعايير المنشورة (عدد الكلمات مثلا)، ومع ذلك فالاختصار يجب أن يكون صفة ثانوية إلى جانب الصفات الأساسية الأخرى كالشمول والدقة، وهناك جوانب أخرى ستؤثر على الاختصار (مستخلص شارح أو إعلامى أو نقدى...).

أما *التكاليف* فهى تتعلق بالمستخلصات على مستويات مختلفة: كالتكاليف الفكرية لإنشاء المستخلص، وتكاليف كل مستخلص فى الدورية المطبوعة، وتكاليف المستخلص فى التوزيع (كجزء من الإحاطة الجارية مثلا)...

والعوامل التى تؤثر على *التكاليف* تختلف من مستوى إلى آخر، فعلى سبيل المثال فطول المستخلص لها تأثير رئيسى على تكاليف النشر المطبوع، ولكن تأثيره أقل بكثير فى قاعدة البيانات الإلكترونية. وتكاليف كتابة المستخلص يمكن أن ينظر إليها من وجهة نظر القائم بالاستخلاص وما يدفع له نظير ذلك...

وهناك اعتبارات أخرى بالنسبة للتكلفة، فاستخدام المستخلصات التى يعدها المؤلف يؤدى إلى تكلفة اقتصادية بالنسبة لمنتجى قواعد البيانات، وقد ترتفع هذه

التكلفة فى المنظور الاجتماعى، حينما يراد لهذا الباحث المؤلف أن ينفق وقته فى البحث والإنتاج بدلا من إعداد المستخلصات.

وعلى كل حال فالتكاليف صفة ذات وجوه متعددة عند ربطها بالمستخلصات والاستخلاص، من اجل ذلك فهى تظهر ضمن جميع المنظورات فى الجدول السابق.

أما *الكثافة Density* فهى مقياس يربط بين صفات الشمول والاختصار، أى أنها تجمع بين منظورى العملية والمنتج، أى أن الكثافة تعبر عن كمية المحتوى المعلوماتى للمستخلص ذى الطول المعين، ويمكن اعتبار الكثافة ذات ارتباط بالانتروبى Entropy (مدى إمكانية تقليل الشك Uncertainty الخاص بالوثيقة الأصلية بالنسبة لقارئ المستخلص). والاختبارات المعيارية الخاصة بتنبؤ صلاحية المستخلصات تتوجه لهذه القضية.

هذا ويعتبر معامل البيانات data coefficient المقترح والمختبر بواسطة الباحث ماتيس (Mathis 1972) مقياسا دقيقا للكثافة، كما سبق شرحه ببعض التفاصيل.

هذا ورضى المستفيد يرتبط بمنظورات العملية والإنتاج التى سبقت مناقشتها خصوصا الدقة والقرائية والشمول، فضلا عن اهتمام منتجى وموزعى المستخلصات بالتكاليف، أى الوصول إلى رضى المستفيد بأقل التكاليف، ومنظور المستفيد بالنسبة للجودة يميل إلى الذاتية والنسبية والديناميكية والمستفيدون عادة يركزون على خدمات المستخلصات والاستخلاص من حيث التكاليف وقيمتها بالنسبة لهم.

رابعاً: اختبار نوعية وجودة المستخلصات فى ثلاث قواعد بيانات:

استخدام الاختبارات الكمية والموضوعية من شأنه أن يزيل الاختلافات المحتملة والتى تنتج عادة من اختلاف وجهات النظر، وبالتالى يزودنا بوجهة نظر عامة عن الجودة والنوعية. وقد تناول تينوبير وزميله (Tenopir, 1993, P. 50) ثلاثة جوانب لاختبارها وهى:

(أ) انتظام الأسلوب وسلاسة القراءة أو القرائية Readability.

(ب) درجة تطبيق وملاحظة المعايير التى وضعتها أنزى (ANSI).

(ج) درجة تمثيلها للمعلومات التى تحتويها الوثيقة Informativeness.

٤-١ سلاسة القراءة:

تتحقق سلاسة القراءة لأى نص مع توفر عدد من العوامل التى تجعل النص مفهوماً، وهذا مقياس نسبى يعتمد على كل من القارئ والنص نفسه من حيث مستوى القارئ الثقافى أو العلمى وذكائه وعمره وهناك بعض المعادلات الخاصة بسلاسة القراءة تم اختبارها بالنسبة لمستوى النص مثلا علامات فليشن Flesch Reading Ease Score

وكشاف فوج Gunnin's Fog Index وفليش كينكاد Flesch- Kincaid Grade Level

وكل واحدة من هذه المقاييس تطبق معادلات معينة على النص لقياس مدى سهولة أو صعوبة قراءة النص. وكل واحدة من هذا المقاييس تستخدم طرقاً مختلفة لحساب الرقم المتوسط للمقاطع اللفظية لكل كلمة، والكلمات فى كل جملة أو نسبة الكلمات لثلاث، مقاطع أو أكثر (Klare,. G. 1963.)

هذا وكشافات الحساب اليدوى للأسلوب وسلاسة القراءة تحتاج لوقت كبير وبالتالى فلا يتم تطبيقها إلا على نطاق محدود، أما الاختبارات الثلاثة السابق الإشارة إليها فهى متاحة على نطاق واسع فى برامج أسلوب ونمو برنامج مثل جراماتيك Grammatik المستخدم فى تحليل النصوص المقروؤة آلياً. كما أن هذا البرنامج الخاص بتحليل الأسلوب يتيح كذلك تحليل سلاسة القراءة والقيام بذلك دون جهد كبير، وذلك على التسجيلات المقتبسة من قواعد البيانات على الخط المباشر أو على الأقراص المكتزة.

وقد قام الباحثان تينوبير وجاكسو (Tenopir, c., 1993) باختيار تسجيلات لاختبارها من قاعدة البيانات بالمؤلف والموضوع وعنوان الدورية وسنة المطبوع وشكل الوثيقة واللغة الأصلية... الخ.

وهذه المعايير قد استخدمت لتعديل محتوى وحجم العينة التى يتم تقييمها. وتوفير عينة ذات حجم مناسب تأخذ وقتاً أطول فى العادة (وتكلفتها عالية فى البيئة الإلكترونية) ولكنه وقت الحاسب الآلى وليس وقت الباحث، وإذا كانت البرنامج يسمح لك بتحميل حقل المستخلصات فقط فيمكن فى نفس الوقت تغذية ملف المستخلصات ببرنامج تحليل الأسلوب Style analyzer software ومعظم نظم الخط المباشر وبرامج الأقراص المكتنزة تسمح للمستفيد بتحديد الحقول التى يمكن تضمينها فى البحث.

وإذا لم يسمح البرنامج بتحميل حقل المستخلصات فقط، فيجب فى هذه الحالة كتابة ماكرو Macro فى برنامج معالجة الكلمات للتخلص من الحقول غير ذات العلاقة من التسجيلات المتضمنة فى الدراسة (مثل الرقم المسلسل/ العنوان/ المؤلف/ الواصفات/ ... الخ). كما ينبغى فحص بناء التسجيلات لكل قاعدة بيانات للتعرف على النموذج المطبق فيها، حتى يمكن الوصول إلى خوارزمية غير غامضة، كما يجب اختبار البرنامج على عدد من التسجيلات للتأكد من أن الماكرو Macro يعمل بالطريقة المطلوبة.

هذا وينبغى تطبيق برنامج معالجة الكلمات أيضاً للتخلص من بعض عناصر البيانات عند تحميل المستخلصات فقط، وذلك لأن بعض هذه العناصر قد يشوه نتائج التحليل، فالمستخلصات قد تحتوى فى قاعدة علم النفس على الملاحظة التالية:

Copyright 1992 American Psychological Ass. all Rights res.

وبالتالى فإن برنامج الماكرو Macro يحب أن يتضمن حذف مثل هذه الجمل. وبعد ذلك يمكن للملف (الذى تم التخلص فيه من الشوائب المذكورة) أن يتم تغذيته بمحلل الأسلوب style analyzer لإتمام عملية تحليل سلاسة القراءة.

ويعتبر كل من المحللين التاليين:Right writes and Grammatik من بين محللى الأسلوب الذين يستخدمان على نطاق واسع، أما شركة ميكروسوفت Microsoft فقد أدخلت ضمن برامجها محلل أسلوب. والهيئات الثلاث تزودنا

بإحصاءات للفقراتParagraph. وللجمل وللكلمات فضلاً عن قياس علامات فليش وكشاف فوج ومستوى فليشن كينكاد السابق الإشارة إليهم. وقد اعتمد تينوبير وجاكسو فى هذه الدراسة التى بين أيدينا على برنامج جراماتيك Grammatik- TV

❖ اختبار سلاسة القراءة: نتائج العينة:

يظهر لنا الجدول (١) التحليل المقارن للمستخلصات من قاعدة بيانات المصادرResource-One data base (RES)

وعددها (١,٠٧٠) مستخلص ومن قاعدة مستخلصات المرشد القرائى.

Reader's Guide Abstracts (RGA).

وعددها (٩٩٩) مستخلص ومن ملخصات مقالات المجلة.

Magazine Articles Summaries (MAS).

وعددها (٩٨٤) مستخلص.

وحتى يمكن إجراء هذا التحليل قام تينوبير Tenopir وزميله باسترجاع تسجيلات من كل قاعدة بيانات بواسطة سؤال موضوعى وكان عن الطائرات كما استخدمت مقاييس فليش كنيكاد وكذلك كشاف جانيجز فوج وذلك لتزويدنا بعلامات Scores تدل على المستويات المعيارية الأمريكية للمستوى (حيث يدل مستوى ١٢ على الطالب فى نهائى المدرسة الثانوية، بينما يدل المستوى ١٧على اجتيازه سنة واحدة فى الدراسة العليا) هذا ويمكن أن يتراوح مقياس فلش بين صفر إلى ١٠٠، وكلما كانت العلامات متدنية كلما كان من الصعب قراءة النص، أما العلامة بين (صفر إلى ٣٠) فتعنى صعب للغاية، وبين (٣٠إلى ٥٠) فتعنى صعب للغاية، وبين (٥٠إلى ٦٠) فتعنى صعب إلى حد ما، (٦٠إلى ٧٠) معيارى. أما التى تكون (أكثر من ٧٠) فتدل على أن الكتابة سهلة إلى حد ما إلى سهلة للغاية عند قراءتها.

هذا والنسبة العالية من المبنى للمجهول Passive وحروف الجر والجمل الكثيرة فى الفقرة Paragrph والكلمات الكثيرة فى الجمل، والمقاطع الكثيرة فى الكلمة، هذا كله من شأنه أن يجعل القراءة أكثر عسراً وصعوبة. وبينما يقوم برنامج محلل الأسلوب بتفسير العلامات تلك، فإن هذه العلامات تكون ذات معنى

أكثر عند مقارنتها بعلامات الكتابة الأخرى المعروفة، ويقدم برنامج جراماتيك Grammatike سلسلة من الرسومات التى تقارن النص الذى يتم تحليله بعلامات ثلاثة كتابات أخرى.

وطبقاً للعلامات التى تم الحصول عليها فإن مستخلصات المصادر تعتبر أكثرها سهولة فى القراءة، إذ حصلت على أعلى علامة فى فليش وهى أقل بمستويين من RGA, MAS ولكن يلاحظ من جانب آخر النسبة المئوية العالية للمبنى للمجهول فى قاعدة المصادر (RES) والتى تجعل قراءة النص أكثر صعوبة. ومع ذلك فقد كان محلل الأسلوب سريعاً ومنتظماً فى حساب العلامات والكشافات، وقد استدعى ذلك تقييم ذاتى من الباحث.

وعلى سبيل المثال فإن علامة طول الجملة لا تدل بالضرورة على خاصية إيجابية لقاعدة بيانات (RES) والسبب فى هذه العلاقة أن الجملة الأخيرة من المستخلص تكون عادة بياناً بنوع المادة المتوافرة (كلمة واحدة أو اثنتين مثل جداول، رسومات) وبالمثل فإن مستخلصات تسجيلات الملخصات Summaries تشمل عادة اسم المؤلف فى الجملة الأخيرة مع وجود الاسم الأول كحرف فقط وبالتالى فإن هذا الإجراء يقلل من متوسط طول الجملة بطريقة جذرية، وهناك مسبب آخر للمتوسط المنخفض إلى حد ما لطول الجملة وهو العدد الكبير من المستخلصات الموجزة فى (MAS). هذا وطول الفقرات فى مستخلصات القارئ Reader's Guide Absthacts يقدم لنا أدنى علامات قرائية ولكن ذلك فى حد ذاته قد يكون أكبر ميزة، حيث يعتبر محلل الأسلوب أن المستخلص يعبر عن فقرة واحدة فى الملف، ومستخلصات القارئ هذه يبلغ طولها مرتين ونصف المرة أطول من المستخلصات فى القاعدتين الأخرتين.

تدلنا العلامات أن مستخلصات المصادر (ERS) هى أسهلها فى القراءة ولكنها تحتوى على أعلى نسبة مئوية من المبنى للمجهول (٣١%) هذا والعلامات ليست معصومة من الخطأ وبالتالى فيجب تفسيرها اعتماداً على الفحص المباشر لمستخلصات كل قاعدة بيانات.

الجدول رقم (١)

تحليل المستخلصات من ثلاث قواعد معلومات ذات اهتمام موضوعى عام

	MAS	RES	RGA
عدد التسجيلات	٩٦٧	١٠٧٠	٨,٣٩
علامات مطلقة			
علامات فليشFlasch Reading Ease	٣٧	٤٦	٣٩
كشاف فوج Gunning's Fog Index	١٧	١٥	١٧
مستوى فليش كينكاد Flesch- Kincaid Grade level	١٣	١١	١٣
متوسط طول الفقرة	١,٩	١,٨	٤,٥
متوسط طول الجملة	١٩,٢	١٧,٦	٢١,٣
متوسط طول الكلمة	٥,٤٣	٥,٠٢	٥,١٨
متوسط المقاطع Syllobler	١,٧٨	١,٦٩	١,٧٣
النسبة المئوية للعلامات			
المبنى للمجهول Passive Voice	٧,٠٩	٣١,٦٣	١٦,٠٤
الحروف	١٣,٩٤	١١,٥٦	١١,٨٥
الجمل القصيرة (١٤ كلمة)	٣٨,٦٠	٤٢,٨١	٢٥,٢٩
١٥- ٣٠ كلمة بالجملة.	٤٣,٥٧	٤٨,٠٣	٥٧,٦١
جمل طويلة (٣٠كلمة).	١٧,٨٢	٩,١٧	١٧,٠٩

٤-٢ درجة الملاءمة مع معايير أنزى (ANSI):

هناك مقياس آخر لنوعية وجودة المستخلصات ودرجة ملاءمتها لمعايير المعهد الأمريكى للمعايير (ANSI) وهو مقياس موضوعى Objective ولكنه يحتاج إلى عمل وحكم إنسانى أكبر. وأحد أجزاء السؤال يمكن الإجابة عليه من معالجات الكلمات أو برامج محلل الأسلوب، ويوصى معهد المعايرة بأن تكون المستخلصات بين مائة إلى ٢٥٠ كلمة، ويبين لنا الشكل رقم (١) متوسط أطوال خمسة قواعد

بيانات تم اختبارها، وهى مستخلصات من المقالات البحثية للدوريات والتى تهدف إلى خدمة الجمهور المهنى وبالتالى فيتوقع أن تكون هذه المستخلصات أطول من المقالات المنشورة فى الدوريات العامة.

هذا يلاحظ أن متوسط كلمات مستخلصات القارئ يصل إلى (١١٠) كلمة أى أنها تقريباً تصل إلى مرتين أطول من ملخصات المقالات Magazine Article وتبلغ ثلاث مرات طول مستخلصات المصادر (RES) وطبقاً لذلك فإن مستخلصات القارئ تعتبر أقربها جميعاً لمعايير أنزى (ANSI) بالنسبة للطول.

٤-٣ درجة الإعلامية: Informativeness:

وهناك جزء آخر من المعايير مرتبطة بالمستخلص الإعلامى وهذه يصعب قياسها، ذلك لأنها تتطلب عينة من التسجيلات من كل قاعدة بيانات ثم الحكم على المستخلص هل هو إعلامى أم شارح أو مزيج بينهما أو غير ذلك من الأشكال، وحتى يمكن إتمام ذلك فلابد من إحاطة الباحث بمعايير أنزى (ANSI) بالنسبة لأنواع المستخلصات.

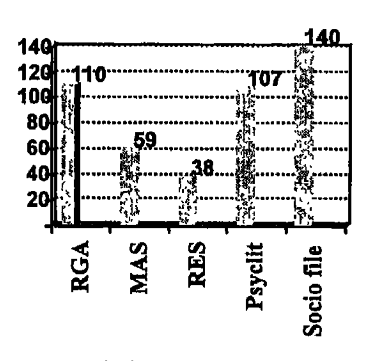
ولقد قام تينوير وزميله بفحص (٣٠٠) تسجيلة مختارة عشوائياً لكل قاعدة بيانات والتى أخذت للتعرف على سلاسة القراءة Readability ويدلنا الشكل (٢) على مقارنة أنواع المستخلصات بقواعد البيانات الثلاثة العامة، ويلاحظ أنه لا يوجد أى مستخلصات نقدية أو مستخلصات الغرض الخاص فى هذه القواعد، واختلفت مع ذلك النسبة المئوية للمستخلصات الإعلامية والشارحة والمزيج.

ودرجة الإعلامية لقاعدة البيانات يمكن قياسها بالنسبة المئوية للمستخلصات الإعلامية المشمولة، كما يمكن قياسها أيضاً بالنسبة المئوية لتوليفة من المستخلصات الإعلامية والشارحة، ولكن عندما تكون النسبة المئوية للمستخلصات الشارحة عالية فى قاعدة البيانات فإن ذلك يعنى شيئين أولهما أنها لا تستجيب لمعايير (ASNI) وثانيهما أنها متدنية من ناحية درجة الإعلامية، وبالتالى فإن المستفيدين من قاعدة البيانات يحتاجون للرجوع إلى المصدر الأصلى بدلاً من الاعتماد على المستخلصات. أما مستخلصات القارئ (GRA) فهى تستجيب لنوعية وجودة المستخلص، ذلك لأن

أكثر من ٨١% (٢٤٥ من ٣٠٠) هى مستخلصات إعلامية وهناك فقط ٩,٧% (٢٩ مستخلص) إعلامى.

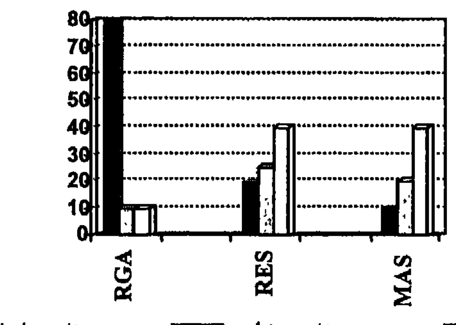
شكل رقم (١)

متوسط أطوال المستخلصات فى خمسة قواعد بيانات بالكلمات



شكل رقم (٢)

أنواع المستخلصات فى كل قاعدة بيانات حسب النسبة المئوية



مستخلص مزيج ▭ مستخلص شارح ▭ مستخلص إعلامى ■

خامساً: اكتشاف المعرفة في قواعد البيانات ونوعية المستخلصات والاستخلاص:

٥-١ تمهيد:

لقد ظهر فى الإنتاج الفكرى خلال السنوات القليلة الماضية عدد من المقالات البحثية عن اكتشاف المعرفة واستخراج البيانات Knowledge Discovery and data Mining (Raafavenetal, 1998) وأن هذا المجال يستخدم طرقا مختلفة لتقييم

البيانات والتعرف على العلاقات المرتبطة بها، والتى يمكن أن تولد معرفة جديدة، وأن هذا المجال(KDD) يتضمن خطوات عديدة تتصل بإعداد البيانات واختيارها والبحث عن نماذج Patterns وتقيم المعرفة وتمحيصها refinement فضلا عن التفسير السليم للنتائج، وهذه الخطوات يتم تكرارها ومراجعتها عدة مرات (Fayyad et al., 1996, P. 41).

وقد قدم كل من بنتو ولانكستر (Pinto, M., Lancaster, F. 1999, 234+) وجهة نظر جديدة عن المستخلصات والاستخلاص تتصل بالأهمية القصوى لنوعية المستخلصات فى أنشطة اكتشاف المعرفة، وأن المستخلصات يجب أن تحتفظ بنوعيات الدقة والقرائية Readability والتماسكcohesion /coherence والاختصار Brevity وتختلف أهمية هذه المعايير Criteria تبعا لمن سيقرأ هذه المستخلصات، فالمستخلصات التى تتوجه أساساً نحو الأغراض البحثية يقل فيها التركيز على معايير القرائية والتماسك.. كما يشير كل من بنتو ولانكستر إلى أن تطبيقات الحاسبات فى تجهيز النص لم تقلل من قيمة المستخلصات، كما أن هذه القيمة لا ينبغى أن تقل مع تطبيق العمليات المعقدة الخاصة باكتشاف المعرفة.

ويمكن فيما يلى تفصيل بعض جوانب دراسة بنتو ولانكستر:

فالمقارنة الواردة فى هذا الصدد تتم بين متطلبات المستخلصات التى تقرأ بواسطة الإنسان عادة بالمتطلبات التى نحتاجها للبحث بالحاسب الآلى، فالإتاحة الواسعة للنص الكامل فى الشكل الإلكترونى فى الوقت الحاضر لا يقلل من قيمة المستخلصات فى أنشطة استرجاع المعلومات حتى مع التطبيقات الأكثر تعقيدا فى اكتشاف المعرفة Knowledge Discovery. هذا والجدوى الاقتصادية اليوم مكفولة ومعترف بها بالنسبة لاختزان كميات ضخمة من النصوص فى شكل يمكن بحثه بالحاسب الآلى، ولكن ذلك لم يقلل من أهمية المستخلصات فمازالت تمثل ملخصات مفيدة يقرأها الإنسان، كما يمكن أخذ المقاييس المعروفة (الاستدعاء والدقة Recall Precision) فى الاعتبار، على الرغم من أن البحث فى النص الكامل سيؤدى فى كثير من الأحيان إلى مستوى غير مقبول من عدم الصلاحية irrelevancy خصوصا وقد أثبتت تينوبير (Tenopir, 1985) منذ فترة أن بحث المستخلصات يمكن أن يكون أكثر فاعلية من البحث فى النص الكامل.

ويلاحظ أن استرجاع المعلومات خصوصا فيما يسمى بأنشطة اكتشاف المعرفة فإن الشمول Exhaustivity والدقة accuracy، تعتبر ذات أهمية بالغة، كما أن المستخلصات التى توجه لأغراض البحث أساسا، لا تعتبر فيها معايير القرائيةReadability والتماسك Cohesion ذات أهمية، هذا بالإضافة إلى أن الاختصارBrevity ليست مرغوبة بالضرورة نظرا لأن استرجاعية المستخلص ستكون مرتبطة بطول المستخلص (أى بعدد نقاط الإتاحة)، كما يعتبر عامل حفظ البيانات الذى اقتراحه ماتيس (Mathis, 1972) مناسبا فى تطبيقات اكتشاف المعرفة نظرا لربطه طول المستخلص باكتمال تغطية المحتوى.

٢-٥ الطرق المعاصرة فى الاستخلاص والمستخلصات:

لقد زاد التجهيز والمعالجة الآلية للنصوص خلال السنوات السابقة مع زيادة قوة التحسيب، وفى نفس الوقت فتكاليف التحسيب والاختزان قد قلت بشكل كبير أى أن النصوص قد أصبحت متاحة فى الشكل الإلكترونى وهى محملة بكمياتها الضخمة على شبكة الإنترنت والشبكة العنكبوتية. www وبالتالى أصبح بحث النص الإلكترونى، هو القاعدة وليس الاستثناء.

ومن أهم التقارير وسلسلة المؤتمرات المتصلة بمعالجة وتجهيز النصوص من اجل استرجاع المعلومات يمكن أن نشير إلى مؤتمرات إسترجاع النص Text Retrieval Conferences (TREC) والذى ينظمه المعهد القومى للمعايير والتكنولوجيا (Pinto, M., 1999, P. 245) وهناك أيضاً مؤتمرات فهم الرسالة(MUC) Message Understanding Conferences والتى تهتم بمعالجة اللغة الطبيعية.

والملاحظة أن الطرق المعاصرة فى معالجة وتجهيز النص لاسترجاع المعلومات تتجاوز مجرد بحث النصوص والتكثيف الآلى والاستخلاص الآلى (ومعظمها يعود إلى الخمسينات من القرن العشرين) اما فى الوقت الحاضر فبالإضافة لهذه البحوث والمعالجات فهناك أنشطة إضافية تتعلق بروابط النص وتوليد النصوص (Text Linkage Text generation) وهى لا تختلف كثيراً من

حيث المبدأ عن الطريقة القديمة وإن كانت قد اتخذت أسماء جديدة مثل (تلخيص النص بدلاً من الاستخلاص والاقتباس، واستخدام فئات النص بدلاً من التكشيف والتصنيف) كما أنها قد تكون أكثر تعقيداً من بعض الوجوه، فضلاً عن استخدام مايسمى بالاستخدام الذكى لمعالجة النصوص وإن كان لانكستر يشك كثيراً فى ان هذه المداخل تعكس الذكاء الحقيقى (Lancaster, F, 1999)



الفصل العاشر

المكانز

الأساسيات والاتجاهات الحديثة والإسهامات العربية

أولاً: تعريف المكنز ووظائفه :

لعل أشهر تعريف للمكنز Thesaurus هو تعريف المنظمة الدولية للتقييس ISO الذى ورد فى مواصفتها القياسية الخاصة بإنشاء وتطوير المكانز أحادية اللغة، والتى تم تعريبها كواصفة قياسية عربية [1] :

يمكن تعريف المكنز من حيث وظيفته أو من حيث بناؤه. فالمكنز من حيث الوظيفة هو أداة لضبط المصطلحات يستخدم للترجمة من اللغة الطبيعية للوثائق أو من لغة المكشفين أو المستفيدين إلى "لغة نظام" أكثر تقييداً (لغة توثيق، لغة معلومات). والمكنز من حيث البناء هو لغة مضبوطة وديناميكية تتكون من المصطلحات المتصلة ببعضها البعض دلالياً وجنسياً (نسبياً) والتى تغطى أحد حقول المعرفة.

والمكنز يصف المحتوى الموضوعى للوثائق .. أما المصطلحات اللازمة لوصف الخصائص الشكلية للوثيقة ـ المصطلحات غير الموضوعية كأسماء المؤلفين أو الهيئات أو الأماكن .. إلخ (الوصف الببليوجرافى) ـ فلا حاجة لأن تكون جزءاً من المكنز على الرغم من أن مثل هذه المصطلحات تعتبر ضرورية لتحديد هوية أو موقع ما تم تكشيفه ويمكن أن تمثل ملحقاً للمكنز.

ويجب أن يعكس المكنز المحتوى الإعلامى للوثائق فى المجموعة التى يطبق عليها، كما يجب أن يحتوى على المصطلحات والإحالات الملائمة للمادة الموضوعية، مع الأخذ فى الاعتبار لكل من لغة مجموعة الوثائق ولغة المستفيدين وحاجاتهم من المعلومات.

---

(١) التوثيق ـ إرشادات لاعداد وتطوير المكانز أحادية اللغة (مواصفة قياسية عربية رقم ٥٧٨-١٩٨٤). المجلة العربية للمعلومات. ـ مج١٣، ع٢ (١٩٩٢). ـ ص ٦٠-٦١.

وعموماً، فالمكنز هو قائمة بالواصفات وعلاقاتها التكافؤية والهرمية والترابطية، ويكون ترتيب وعرض الواصفات وعلاقاتها بما يخدم بكفاية وفعالية في تحليل محتوى أوعية المعلومات واسترجاعها (٢) .

ويختلف المكنز بالطبع عن نظام التصنيف Classification system وعن قائمة رؤوس الموضوعات List of subject heading

وإن بدأت قوائم رؤوس الموضوعات التقليدية تقترب من نظم المكانز، فإن قائمة مكتبة الكونجرس لرؤوس الموضوعات Library of Congress Subject Headings طبقت نظام الإحالات المستخدم في المكانز إبتداء من الطبعة الصادرة عام ١٩٨٩، وقد فعلت ذلك قائمة سيرز لرؤوس الموضوعات Sears List of Subject Headings في طبعتها الأخيرة.

ويعتبر المكنز من أهم أنواع لغات التكشيف، وهو وسيلة لوصف محتوى الوثائق ومن ثم يستخدم في عملية إدخال المعلومات وكذلك في عملية ترجمة أو تحويل استفسارات المستفيدين إلى لغة النظام وهو بذلك حجر الزاوية في نظم استرجاع المعلومات.

إن الوظائف الأساسية للمكنز هي إنه يتيح للمكشف تمثيل المادة الموضوعية المحتواة في الوثائق بطريقة ثابتة وموحدة، كما أنه يحضر المصطلحات المستخدمة من جانب الباحث في توافق مع المصطلحات المستخدمة من جانب المكشف. وهو بالإضافة إلى ذلك يمد بالوسائل التي تمكن الباحث من أن يعدل استراتيجية البحث من أجل تحقيق استدعاء عال أو إحكام عال كما تتطلب الظروف المتنوعة (٣) .

وعلى ذلك فالمكنز هو أداة المكشف وهو أيضاً أداة الباحث، وكلاهما مستفيد منه، فالمكشف يعتمد عليه في الحصول على الواصفات الملائمة التي يستخدمها في وصف محتويات الوثائق. والباحث يعتمد عليه أيضاً في الحصول على الواصفات

(٢) محمد فتحي عبد الهادي. المكانز كأدوات للتكشيف واسترجاع المعلومات. ـــ القاهرة: مكتبة غريب، ١٩٨٩. ـــ ص ٢٨

(3) Lancaster, F.W. Vocabulary control for information retrieval. – Washington, D.C: Information Resources Press, 1972. – p. 185.

المناسبة التى يستخدمها فى وصف حاجاته، وهى تلك التى تتفق مع واصفات النظام. فالمكنز إذن حلقة الوصل بين المكشف والباحث وهو أيضاً اللغة المشتركة بينهما.

أن اثنين من المكشفين سوف يتفقان على المصطلح أو المصطلحات اللازمة لوصف رأس معين، إذا اختيرت المصطلحات من قائمة سابقة الإنشاء. وعلاوة على هذا فإنه من الأفضل أن تتاح أمام الباحث قائمة بمصطلحات النظام يختار منها ما يفيده فى صياغة استفساره أو طلبه.

وهكذا فإن الغرض الأولى للمكنز هو استرجاع المعلومات، والذى يمكن تحقيقه باستخدام المكنز فى التكشيف (فكرياً أو آلياً) لقاعدة بيانات و/أو فى بحثها.

وتتضمن الأغراض الثانوية للمكنز المساعدة فى الفهم أو الإدراك العام للمجال الموضوعى، وتقديم "خرائط دلالية" بإظهار العلاقات الداخلية بين المفاهيم، وأيضاً المساعدة فى تقديم تعريفات للمصطلحات. ومن التطبيقات الأخرى توليد قوائم كلمات مفتاحية يمكن أن تشكل الأساس للتخطيط وغير ذلك من الأعمال. وهو أيضاً يدعم الاستخلاص بمساعدة الحاسوب (٤) .

وعادة ما يغطى المكنز مجالاً موضوعاً محدداً سواء أكان هذا المجال واسعاً أو ضيقاً. ويمكن أن يكون المكنز خاصاً بنظام معلومات مؤسسة ما مثل مكنز اليونسكو Unesco Thesaurus الذى يغطى قطاعات التربية والثقافة والعلوم والإتصال والمعلومات وهى تلك التى تهتم بها منظمة اليونسكو فى أنشطتها وتعاملاتها. وقد نشأت فى الآونة الأخيرة بعض المكانز التى تغطى مجالات موضوعية متنوعة مثل المكنز الموسع" الصادر عن مؤسسة عبد الحميد شومان بعمان ومركز جمعة الماجد للتراث والثقافة بدبى.

وقد يكون المكنز أحادى اللغة أو متعدد اللغات والمكنز أحادى اللغة هو الذى يشتمل على المصطلحات فى لغة واحدة فقط، أما المكنز متعدد اللغات فهو الذى يستخدم للتكشيف والبحث فى عدة لغات مثل الإنجليزية والفرنسية والعربية، ومن ثم يشتمل المكنز على المصطلحات فى لغة ما ومقابلاتها فى اللغات الأخرى. ومن

(4) Aitchison, Jean. Thesaurus consrtuction and use: a practical manual/ Jean Aitchison, Alan gilchrist, David Bawden. – 3rd ed. – London: Aslib, 1997. – p.1

أمثلة المكانز متعددة اللغات مكنز "الجامعة" الصادر عن مركز التوثيق والمعلومات بالأمانة العامة لجامعة الدول العربية.

وفيما يتعلق بالشكل المادي للمكنز، فإن المكنز قد يكون فى شكل مطبوع على هيئة كتاب ينشر ويتاح للتوزيع على نطاق واسع، أو يطبع ويتاح على نطاق ضيق داخل المؤسسة التى تستخدمه.

ويمكن أن يتاح المكنز فى شكل الكترونى، أى يُستخدم من خلال شاشة حاسوب. وميزة مثل هذا العرض أنه يمكن أن يقدم بيانات كاملة عن كل مصطلح، فإن تسجيلة المصطلح يمكن أن تشمل ــ إضافة إلى التبصرات التوضيحية والعلاقات المعيارية ــ بيانات عن نوع التسجيلة، وملاحظات المفهرس، وتاريخ الإنشاء، وبيانات عن آخر تحديث لها. (٥)

ثانياً : بناء المكنز :

يتطلب تحقيق المكنز للأغراض المشار إليها فيما سبق أن يشتمل المكنز على المصطلحات المقننة الصالحة للاستخدام فى نظام المعلومات، وأن يعرض العلاقات المختلفة بين هذه المصطلحات. وهذا ما يعرف ببناء المكنز.

٢/١ ضبط المصطلحات :

يشتمل المكنز فى العادة على نوعين رئيسيين من المداخل هما: الواصفات واللاواصفات. أما الواصف Descriptor أو المصطلح المفضل فهو مصطلح مقنن يستخدم للتعبير عن أو التمثيل الواضح للمفاهيم أو المادة الموضوعية فى الوثائق واستفسارات الباحثين.

والواصفات قد تدل على المفاهيم أو تركيبات المفاهيم.

كما قد تدل على كيانات فردية، وتسمى هذه المصطلحات أيضاً أسماء الأعلام أو الهويات، ومنها: أسماء مشروعات، أسماء جغرافية، أسماء أشخاص أو هيئات، أسماء تجارية، أسماء أعمال فنية ....إلخ.

---

(5) Ibid. p. 93.

أما اللاواصف Non- Descriptor فهو المصطلح غير المسموح باستخدامه فى التكشيف. واللاواصفات تشمل المترادفات والأشكال الأخرى من المصطلحات المفضلة أو المجازة. ومثل هذه المصطلحات يحال منها إلى المصطلحات المجازة أو الواصفات.

ويجب أن يصاغ كل واصف فى المكنز بطريقة تجعله يحمل المعنى المقصود بوضوح حتى يفهمه المستفيد وحتى يصل إليه بسرعة.

وقد يكون الواصف كلمة واحدة، وقد يتكون من كلمتين أو أكثر. ومن أشكال الواصفات: الكلمة الواحدة مثل "الزراعة"، الصفة والموصوف مثل "التخطيط الاقتصادى"، المضاف والمضاف إليه مثل "مسك الدفاتر"، الربط بحروف الجر (فى حالات قليلة) مثل "التعليم بالمراسلة".

ويجب أن يكون المصطلح فى شكل اسم، سواء استخدمنا الأسماء المفردة (كلمات واحدة) أو الجمل الأسمية. ويجب تحاشى صيغة الفعل واستخدام صيغة المصدر بدلا منها مثل:

الاتصال	**بدلا من**	**يوصل**
الإدارة	**بدلا من**	**يدير**
التنمية	**بدلا من**	**ينمى**

ويمكن استخدام شكل الجمع عندما يكون المصطلح هو اسم عد، أى الاسم الذى يمكن الإجابة عليه بالسؤال "كم عدد؟"، واستخدام شكل المفرد عندما يكون المصطلح هو اسم جمعى Mass noun أى الاسم الذى يعبر عن "ما مقدار؟". وعادة ما يستخدم شكل المفرد للعمليات المحددة مثل: الصيانة، والخواص المميزة مثل: الذوبان، والأشياء الفريدة مثل: الأوكسجين. ونحتاج إلى صيغة المثنى بالإضافة إلى صيغة المفرد وصيغة الجمع بالنسبة للغة العربية عندما يكون أصل الموضوع من الأسماء الزوجية مثل: "الرئتان".

وهناك بعض المصطلحات التى تحتاج إلى نوع من الإيضاح أو التفسير، مثل المصطلحات التى لها أكثر من معنى مقبول، أو المصطلحات التى تستخدم فى معنى خاص إلى حد ما.

ويمكن توضيح معانى مثل هذه المصطلحات وبيان نطاقها وفق طريقة من الطرق التالية:

(أ) التعبيرات بين أقواس مثل: الخلايا (أحياء)

الخلايا (كهرباء)

(ب) التبصرات التوضيحية والتعريفات. وهى تفسيرات أو شروح قصيرة تعطى عند الحاجة، لتفادى الغموض الدلالى للمصطلح ولتأكيد الاستخدام الصحيح له داخل سياق المكنز. والتبصرات أو التعريفات تصحب الواصف فى القسم الرئيسى من المكنز ولكنها لا تشكل جزءاً منه بينما تعتبر التعبيرات بين أقواس بعد المصطلح جزءاً منه.

تبقى الإشارة إلى أن المفاهيم متعددة الكلمات تدخل فى المكنز وفقاً لطريقة من الطريقتين التاليتين:

ـ فى معظم الحالات كمفاهيم مسبقة الربط، أى أن يوضع الواصف فى شكله متعدد الكلمات. مثل : خطط التنمية

ـ فى عدد قليل من الحالات كمفاهيم لاحقة الربط، أى أن يركب المفهوم بالربط بين واصفين أو ثلاث من الواصفات مثل [٦] :

Societal assessment USE Evaluation AND Societal Criteria

وهكذا فإن المصطلحات يمكن أن تربط قبل أن تدخل النظام ومن ثم فسوف تسترجع كما هى، أو تربط خلال الاسترجاع لتمثيل المفهوم كما فى المثال السابق.

٢/٢ العلاقات بين المصطلحات فى المكانز:

يعتبر تقديم العلاقات المتبادلة بين المصطلحات التى يشتمل عليها المكنز سمة أساسية من أهم السمات المرتبطة به.

(6) Spines thesaurus, vol. 1: Rules, conventions, directions for use. – Paris: Unesco, 1976. – p. 38-39.

وهناك ثلاثة أنواع من العلاقات هى [٧] :

علاقة التكافؤ، العلاقة الهرمية، علاقة الترابط.

(١) علاقة التكافؤ أو التساوى Equivalence Relation :

هى العلاقة بين المصطلحات المفضلة والمصطلحات غير المفضلة، حيث نجد أن هناك مصطلحان (أو أكثر) يشيران إلى نفس المفهوم. ومعنى ذلك أنه توجد بعض المفاهيم التى يمكن التعبير عنها بأكثر من تسمية واحدة. ويمكن النظر إلى هذه التسميات المتعددة على أنها متساوية، أو متساوية تقريباً فى الدلالة على تلك المفاهيم. ومن ثم يمكن استخدام تسمية واحدة فقط من بين التسميات المتعددة ـــ وهى المفضلة فى العادة ـــ لاسترجاع الوثائق المتعلقة بالمفهوم.

ويوجد نوعان من الإحالات: إحالة استخدم، وإحالة مستخدم لـــ.

أما إحالة استخدم (اس) Use فهى التى تقود من الواصفات أو من المصطلحات غير المفضلة إلى الواصف أو المصطلح المفضل، ومن ثم فهى تفيد فى توجيه المستفيد إلى الواصف المناسب فى المكنز. ومن أمثلتها:

الهاتف	اس التليفون
القرصنة الجوية	اس خطف الطائرات
دور الكتب	اس المكتبات
البدانة	اس السمنة

والإحالة العكسية أو المتبادلة لإحالة استخدم السابقة هى إحالة مستخدم لـــ(س ل) Used for وهى تصحب المصطلح المفضل الذى تحيل إليه إحالة استخدم

ومن أمثلتها : السمنة

س ل البدانة

(٢) العلاقة الهرمية : Hierarchical Relation

(٧) انظر:

Aitchison, Jean. Thesaurus construction and use. 1997.-p. 47-64.

هى العلاقة التى تعبر عن علاقة العلوية (الوضع فى مرتبة أعلى) Super ordination والتابعية (الوضع فى مرتبة أدنى) Subordination للمفاهيم. ومن أنواعها: علاقة الشمول، وعلاقة الجزء/ كل.

وفى علاقة الشمول نجد أن المصطلح الشامل (العريض) يمثل طائفة مفاهيم، أما المفهوم المعبر عنه بواسطة مصطلح مخصص (ضيق) فإنه دائماً عضو من أعضائها، ويختلف المفهوم المخصص عن المفهوم العام فى خاصية واحدة على الأقل.

ومن الأمثلة :

القيمة

م ض القيمة الاجتماعية

القيمة الأخـــلاقية

القيمة الاقتصادية

القيمة الثقافية

وفى علاقة الجزء/ كل نجد أن مصطلح الكيان entity (المصطلح العريض) يمثل طائفة أشياء أو مفاهيم، أما الشىء أو المفهوم المعبر عنه بواسطة مصطلح ضيق فإنه يمثل أحد أجزائها، وعلى ذلك فالمصطلح المخصص هنا جزء من المصطلح العام. ومن الأمثلة:

الجيولوجيا

م ض الجيولوجيا الاقتصادية

الجيولوجيا الهندسية

England

NT East Anglia

وتمثل العلاقة الهرمية فى معظم المكانز بواسطة إحالات المصطلح الأعرض (الأوسع) (م ع) مشيرة إلى علاقة المفهوم من حيث كونه أعلى فى الرتبة، والمصطلح الأضيق (م ض) مشيرة إلى العلاقة العكسية أو المتبادلة.

مثال : النحاس المعادن

م ع المعادن م ض النحاس

ج) علاقة الترابط أو الاقتران Associative Relation

وهى تستخدم في العادة لتغطية العلاقات الأخرى بين المفاهيم المتصلة ببعضها البعض اتصالا وثيقاً غير علاقة الاتصال الهرمى. وعادة ما يشار إلى علاقة الترابط بواسطة الإحالة الخاصة بالمصطلحات المتصلة أوالمرتبطة (م ت). وهذه الإحالة تذكر المستفيد عند فحصه للمصطلح (أ) بوجود المصطلح (ب) والذى قد يكون أكثر ملاءمة من المصطلح (أ) فى تشخيص مفهوم فى وثيقة أو استفسار لأحد الباحثين. ويجب أن تعد علاقة الترابط المتبادلة، أى أن تكون الإحالة (أ) متصلة بـــ (ب) والعكس أى (ب) متصلة بـــ(أ). ومن أمثلتها:

التدريس

م ت معينات التدريس

معينات التدريس

م ت التدريس

المحاسبة

م ت المحاسبون

البحث العلمى

م ت الاختراعات العلمية

العنف

م ت ضحايا العنف

تحليل البيانات

م ت البيانات

التدريس

م ت الطلاب

ويوضح المثال التالى (المأخوذ من مكنز الفيصل) شبكة العلاقات الخاصة بأحد الواصفات

الاحرام

س ل مناسك الإحرام

م ض سنن الإحرام

محظورات الإحرام

م ع أركان الحج

أركان العمرة

م ت مواقيت الحج

ثالثاً: عرض المصطلحات بالمكانز:

هناك العديد من الطرق التى تعرض بها المصطلحات فى المكانز سواء فى شكل ورقى أو فى شكل الكترونى. وعموماً فإنه يمكن تقسيم هذه الطرق إلى أربعة أنواع رئيسية هى:

(أ) العرض الهجائى الذى يُظهر بالنسبة لكل مصطلح التبصرات التوضيحية وعلاقات التكافؤ والعلاقات الهرمية والترابطية.

(ب) العروض الهرمية التى تُولد من العرض الهجائى.

(ج) العروض النسقية أو المنهجية والهرمية التى تظهر البناء الكلى للمكنز وكل مستويات الهرمية.

(د) عروض الرسومات أو العروض البيانية بأنواعها المختلفة.

وتشمل كل المكانز العرض الهجائى والذى قد يُصحب أو لا يُصحب بالأشكال الأخرى من العروض. ونتناول هذه العروض بإيجاز فيما يلى:

العرض الهجائى:

عادة ما يشتمل هذه العرض على المصطلحات المفضلة والمصطلحات غير المفضلة فى ترتيب هجائى واحد.

وتختلف كمية البيانات المرتبطة بالمصطلح حسب طبيعته وعلاقاته بالمصطلحات الأخرى، كما تختلف نوعية البيانات من مكنز لآخر، وهناك أيضاً

بعض الاختلافات فيما يتعلق بترتيب البيانات وما يتعلق بالرموز الدالة على علاقات المصطلح

فالرموز الشائعة هى:

المصطلح المفضل

ت و SN تبصرات توضيحية أو تعريفات

س ل UF إحالات للمصطلحات غير المفضلة

م ع BT إحالات للمصطلحات الأعرض

م ض NT إحالات للمصطلحات الأضيق

م ت RT إحالات للمصطلحات المرتبطة

المصطلح غير المفضل

ا س Use المصطلح المفضل

وهناك نظام أخر للرموز على النحو التالى :

المصطلح المفضل

+ المصطلحات غير المفضلة أو المصطلحات المساوية

> المصطلحات الأعرض

< المصطلحات الأضيق

– المصطلحات المرتبطة

المصطلح غير المفضل

المصطلح المفضل

وهناك بعض المكانز التى تعطى مصطلحات أعرض أو أضيق لأكثر من مستوى هرمى واحد فى العرض الهجائى

مثال : الأسر المحرومة

م ع ١ الأسر

م ع ٢ المؤسسات الاجتماعية

م ض ١ الأسر الشريدة

م ض ١ أسر العائل الواحد

وتشتمل بعض المكانز على كشافات مستقلة يطلق عليها الكشافات البرميوتية أو كشافات التباديل Permuted Index وقد يكون الكشاف من نمط كشاف الكلمات الدالة فى السياق KWIC أو من نمط كشاف الكلمات الدالة خارج السياق KWOC .

نموذج (١) كشاف تباديلى من مكنز الفيصل

ثقيف
ثقيف ٠٥,٤٠س
دعوة ثقيف ٢٥,٧٠,١٠س
ثلاث
الطلاق ثلاثا ١٤,٢٥,١٠ع
ثلاثة
المساجد الثلاثة ٨٩ط
ثلاثيات
ثلاثيات أحمد ٦٣,٣٠,١٠د
ثلاثيات البخاري ٥٧,١٠,٢٠د
ثمار
بيع الثمار ١٠,٠٢,٤٠ع
زكاة الثمار ٧٦,٥٥ط
زكاة الثمار الموقوفة ٧٦,٧٠ط
زكاة الزروع والثمار ٧٦,٢٥/٧٠ط
نصاب الثمار ٧٦,٥٥,٤٠ط
ثواب
الثواب ٤٠ق
ثوب
الصلاة في ثوب مغصوب ٥٥,٦٧ط
ثور
غار ثور ٨٨,١٠,٣٠ط
مذهب أبي ثور ٠٧,٢٥ط
ثوري
مذهب الثوري ٠٧,١٥ط
ثويبة
ثويبة ١٥,٢٠س
ثياب
كف الثياب ٥٦,٤٠,٣٠ط
جاثية
سورة الجاثية ١١,٧٠,٥٠ح
جار
حقوق الجار ٢٦,٤٠ع

جامع
تاريخ جامع عمرو ٢٥,٤٠ب
جامع الأصول ٧٢,٢٠د
جامع الصحيحين ٧٢,١٠د
الجامع الصغير ٧٢,٢٠د
جامع عمرو ٢٥,٢٠ب
صحيح الجامع الصغير ٨٥,٥٠د
جاهلي
عصر جاهلي ٠٥س
جاهلية
أديان الجاهلية ٠٥,٠٥س
أنكحة الجاهلية ١٢,٣٤ع
الرق في الجاهلية ٠٥,٦٥س
عادات الجاهلية ٠٥,١٥س
مجتمع الجاهلية ٠٥,١٠/٧٥س
المرأة في الجاهلية ٠٥,٧٠س
جبرية
جبرية ٨٦ق
عقود جبرية ٠٥,١٢,١٠ع
جبريل
جبريل عليه السلام ٢٢,١٠ق
جبل
جبل أحد ٨٨,٣٠,٢٠ط
جبل الرحمة ٨٥,٧٠,١٠ط
جبيرة
شروط مسح الجبيرة ٣٦,١٠ط
صفة مسح الجبيرة ٣٦,٢٠ط
مبطلات مسح الجبيرة ٣٦,٢٠ط
مسح الجبيرة ٣٦ط
جحش
السيدة زينب بنت جحش ٨٠,٦٠س
فضائل السيدة زينب بنت جحش ٨٠,٦٠,١٠س
مرويات زينب بنت جحش ٦٤,٢٠,٣٠د

وعموماً فإن مثل هذا النوع من الكشافات يعتمد على استخدام كل كلمة مهمة من كلمات المصطلح متعدد الكلمات ككلمة مدخل فى الترتيب الهجائى (انظر شكل ١)

ومن الطرق الجديدة ما اتبعه مكنز اليونسكو فى طبعته الثانية Unesco Thesaurus (٨) حيث تم دمج القائمة الهجائية التى ترتب الواصفات واللاواصفات مباشرة تحت كلماتها الأولى مع القائمة التبادلية التى ترتب بالكلمات المهمة فى الواصفات أو اللاواصفات المركبة (غير الكلمة الأولى فى السياق) فى ترتيب هجائى واحد وذلك تيسيراً على المستفيد وتوحيداً للاستخدام.

العروض الهرمية التى تُولّد من العرض الهجائى

عندما يكون المكنز فى شكل مقروء آلياً فإنه من السهل إنتاج عروض هرمية من معلومات المصطلحات الأعرض/ الأضيق المقدمة فى المكنز الهجائى. ومن المعروف أنه من غير الممكن عرض كل مستويات الهرمية فى وقت واحد فى الترتيب الهجائى، وحتى إذا تم ذلك فإنه ليس من السهل التمييز بينها. لكن العرض الهرمى يتغلب على هذه الصعوبة حيث أنه ينتج شجرات هرمية Hierarchical Trees مرتبة بالمفاهيم العريضة على رأس الشجرة، وتحت كل منها التفريعات فى ترتيب هجائى، وتحت كل تفريع تفريعاته فى ترتيب هجائى أيضاً ... وهكذا.

ومن ثم فهى تجعل الفحص الموضوعى أكثر سهولة (أنظر شكل ٢)

العروض النسقية أو المنهجية:

إن العرض النسقى systematic (يعرف أيضاً بالعرض المصنف أو الموضوعى أو الهرمى) يرتب فئات أو هرميات المصطلحات وفقاً لمعانيها وعلاقاتها الداخلية Inter-relationships المنطقية.

ويشتمل المكنز النسقى على قسمين:

- القسم النسقى أو المنهجى (انظر شكل ٣)
- القسم الهجائى (سواء ككشاف أو كمكنز كامل يوجه المستفيد للأجزاء الملائمة فى

(٨) إبراهيم عبد الموجود حسن. الاتجاهات الحديثة فى بناء وعرض المصطلحات فى المكانز: مكنز اليونسكو نموذجاً. ـــ دراسات عربية فى المكتبات وعلم المعلومات. ـــ س٣، ع١ (يناير ١٩٩٨). ـــ ص ٢٢، ٢٣.

القسم المنهجى).

وعادة ما يكون الرابط بين القسمين هو رمز معين يضاف فى القسم المنهجى ويعمل كإحالة فى المكنز الهجائى.

ومن نماذج هذه العروض:

شكل (٢) العروض الهرمية التى تُولد من العرض الهجائى

INSPEC thesaurus

Hierarchies

telecontrol
. telerobotics

temperature
. atmospheric temperature
. boiling point
. Debye temperature
. ferroelectric Curie temperature
. magnetic transition temperature
. . Curie temperature
. . Morin temperature
. . Neel temperature
. melting point
. plasma temperature
. superconducting transition temperature

test equipment
→ . automatic test equipment
. battery testers

testing
. aerospace testing
. . aircraft testing
. antenna testing
→ . automatic testing
. . automatic test software
. boundary scan testing
. built-in self test
. cable testing
. . power cable testing
. computer testing
. conformance testing
. electron beam testing
. electron device testing
. . electron tube testing
. . semiconductor device testing
. . superconducting device testing
. electronic equipment testing
. . circuit testing
. . . integrated circuit testing
. . . printed circuit testing
. . computer equipment testing
. . telecommunication equipment testing
. environmental testing
. . environmental stress screening
. impulse testing
. insulation testing
. . insulator testing
. life testing
. logic testing

Top term entry to hierarchies

automatic teller machines
BT EFTS
TT computer applications
data handling
finance
RT bank data processing
banking
computer networks
point of sale systems
CC C5540; C7120; D2020E
DI January 1985
PT EFTS

→ **automatic test equipment**
UF computerised test equipment
BT computerised instrumentation
test equipment
TT computer applications
instrumentation
→ test equipment
RT automatic test software
automatic testing
logic analysers
portable instruments
quality control
reliability
testing
CC B7210B; C3200; C3380; C7410H
DI January 1969

automatic test software
includes software for computerised instrumentation
BT automatic testing
computer software
electrical engineering computing
TT computer applications
computer software
engineering
testing
RT automatic test equipment
computerised instrumentation
electronic engineering computing
CC B7210B; C7410H
DI January 1995
PT automatic test equipment
automatic testing

→ **automatic testing**
UF ATPG
self testing
NT automatic test software
→ BT testing
TT testing
RT automatic optical inspection

automatic test equipment
boundary scan testing
built-in self test
design for testability
quality control
reliability
CC B7210B; C3380; C7410H
DI January 1969

automation
NT branch automation
factory

شكل (٣) القسم المصنف فى مكنز الفيصل

٥٠ب ...مبادىء الإسلام

* ....عموم الرسالة (انظر ٩٢. ٧٠س)

١٠ب ....الدعوة الإسلامية

١٠. ١٥ب ....دعاة

١٠. ١٥. ١٠ب ....أخلاق الدعاة

١٠. ١٥. ٢٠ب ....إعداد الدعاة

١٠. ١٥. ٣٠ب ....صفات الداعية

١٠. ١٥. ٤٠ب ....واجبات الداعية

١٠. ١٧ب ....فقه الدعوة

١٠. ٢٠/٢٩ب ....منظمات إسلامية

١٠. ٢٢ب ....منظمة المؤتمر الإسلامي

١٠. ٢٢. ٣٠ب ....البنك الإسلامي للتنمية

١٠. ٢٢. ٤٠ب ....صندوق التضامن الإسلامي

١٠. ٢٥ب ....رابطة العالم الإسلامي

١٠. ٢٧ب ....الندوة العالمية للشباب الإسلامي

١٠. ٢٨ب ....منظمة الدعوة الإسلامية

١٠. ٣٠/٥٠ب ....جماعات إسلامية

* ....مهدية (انظر ٦٥. ٦٠ف)

١٠. ٣٢ب ....الإخوان المسلمون

١٠. ٣٥ب ....الجماعة الإسلامية

١٠. ٣٧ب ....جماعة التبليغ

١٠. ٣٨ب ....أنصار السنة المحمدية

١٠. ٣٩ب ....الجمعية الشرعية

١٠. ٤٠ب ....جماعة التكفير والهجرة

١٠. ٤٢ب ....جماعة النور

١٠. ٤٥ب ....جمعية علماء الجزائر

١٠. ٤٧ب ....حزب التحرير الإسلامي

١٠. ٥٥/٧٠ب ....دعوات إصلاحية

١٠. ٥٧ب ....الدعوة السنوسية

١٠. ٦٠ب ....دعوة الشيخ محمد بن عبدالوهاب

١٠. ٦٢ب ....دعوة الشيخ محمد عبده

١٠. ٦٥ب ....دعوة جمال الدين الأفغاني

(أ) المجموعـــات الموضوعية العريضة: ترتب المصطلحات فى هذا النوع ترتيباً هجائـــياً تحت مجموعات موضوعية عريضة. ويظهر الرقم الكودى الخاص بالمجموعـــة الموضـــوعية أمام المصطلح الذى ينتمى إلى هذه المجموعة فى القسم الرئيسى الهجائى.

ومن الأشكال الجديدة فى هذه الفئة ما قدمه مكنز اليونسكو Unesco Thesaurus فى طبعته الثانية حيث أن العرض الهجائى يشكل القسم الرئيسى من المكنز وهو يتضمن الهرميات متعددة المستوى. أما العرض المنهجى وهو فى وضع إضافى أو مساعد فإنه يطلق عليه المكانز المصغرة Micro thesauri . وتتضمن القائمة الهرمية بالمكانز المصغرة صفحة واحدة لكل مكنز مصغر على حدة ويندرج بدوره مع المكانز الأخرى تحت الحقل أو الميدان المعرفى الرئيسى ويتم التعرف عليه من خـــلال رقمـــه واسمه، وترتب القائمـــة أولاً حسب الحقول المعرفيـــة: ١- التربية ٢- العلوم ٣- الثقافة إلخ ومن ثم بتسلسل أرقام المكانز المصغرة ويندرج تحت المكنز المصغر كل واصفات المصطلحات الأعلى وتفريعاتها الأضيق المخصصة من الواصفات.

وترتب المداخل هجائياً وتحت كل مدخل واصف نجد:

- اللاواصف أو اللاواصفات
- التنازل الهرمى للواصفات، وترتيب الواصفات المخصصة تبعا

لتنازل المستويات الهرمية وضمن كل مستوى هرمى فى ترتيب هجائى واحد [٩].

(ب) البناءات الشجرية:

من أبرز نماذجها ما ورد فى قائمة رؤوس الموضوعات الطبية Medical Subject Headings الصادرة عن المكتبة القومية للطب بالولايات المتحدة حيث رتبت المصطلحات فى مجموعات موضوعية وداخل كل مجموعة رتبت المصطلحات ترتيباً هرمياً والرابط بين القائمة الهجائية وموقع المصطلح فى هذا النظام هو رمز مفصل.

(٩) المصدر السابق. ص ٢٥-٢٧.

(ج) العرض الوجهى:

توجد بعض المكانز التى تبنى باستخدام طريقة التحليل الوجهى ومن ثم فإنها يمكن أن تشتمل على عرض وجهى للمصطلحات. وهناك نوعان: المجموعات الوجهية العريضة التى استخدمت أثناء تجميع المكنز والتى يمكن أن تأتى كمحلق للقسم الهجائى. وعادة ما ترتب المصطلحات هجائياً تحت رؤوس الأوجه. أما النوع الثانى وهو المهم هنا فإنه يتطلب تكامل تصنيف وجهى مفصل مع المكنز كما فى حالة المكنز الوجهى Thesauro-facet وحيث ينشأ التصنيف الوجهى والمكنز الهجائى فى نفس الوقت أثناء عملية التجميع (انظر شكل ٤).

عروض الرسومات أو العروض البيانية

إن مصطلحات التكشيف وعلاقاتها الداخلية يمكن عرضها فى شكل ثنائى البعد حيث يمكن للمستفيد أن يختار منه تلك المصطلحات المناسبة له. إن عروض الرسومات البيانية تساعد فى تحسين الوصول كما أنها يمكن أن تكون أداة نافعة أثناء بناء المكنز فى إحضار المصطلحات المتصلة معاً فى عناقيد أو مجموعات دلالية Semantic clusters .

وهناك عدة أنواع من عروض الرسومات أبرزها:

بناء شجرة العائلة Family tree structures

وفى هذا النوع يوضع المصطلح العريض على الرأس، بينما تطبع المصطلحات الأضيق فى مواضع فرعية، ويشار إلى العلاقة بينهما بواسطة الخطوط الرابطة الرأسية. وعادة ما تعطى رسمة شجرة العائلة رمزاً يستخدم كعنوان لكل المصطلحات فى الرسمة.

الأسهم:

يتكون النظام هنا من ترتيب للواصفات فى مجموعات دلالية بتخصيص فرخ مشبك وإعطاء مواضع ثابتة لكل واصف بالنظر إلى المحاور الأفقية والرأسية ومن ثم تحدد الاحداثيات. ويمكن إظهار العلاقات بين الواصفات بواسطة: "خرائط ربط" أو "أسهم" أو "خرائط مصطلحات" تظهر العلاقة بين المصطلحات. ويشتق رمز المصطلح من البناء الشبكى للسهم أو الخريطة ويعرض الرقم الكودى أمام المصطلح

فى القائمة الهجائية. (انظر شكل ٥)

وعموماً فإن عرض الرسومات البيانية مثل التصنيف الوجهى يحضر المصطلحات المتصلة ببعضها البعض فى تقارب مادى ويتيح للمكشف والباحث رؤية كل العلاقات بنظرة.

شكل (٤) مكنز وجهى

Root thesaurus

Systematic display

K Electrotechnology

KB/KO Electrical engineering (continued)
KE/KJ Electrical equipment (continued)
KIP Electrical protection equipment (continued)
KIP.V Electric contact protection
 '-Electric contacts KNR

 (By construction)

KIP.W Double electrical insulation
 '-Electrical insulation CYB.K
 '-Electrical insulation devices KNX

 (By connection to earth)

KIP.X Earthing
 =Earth (electric)
 =Earthing systems
 =Electric grounding
 =Grounding (electric)
 '-Earthing reactors KHC.E
KIP.XE Earth electrodes
 '-Electrodes KNW
KIP.XH Earth conductors
 =Protective conductors
 '<Electric conductors KNN
KIP.XN Earthing switches
 =Automatic earthing switches
 '<Switches KJH
KIP.XR Neutral conductors
 '<Electric conductors KNN

KJ Switchgear
 '>Fuses KIP.M
 '-Bus-bars KNN.B
 '-Electric control equipment KIB
 '-Switching substations KDS SH

Alphabetical display

Cinematography LPM
 < Photography
 - Sprockets (cinematography)
 '- Film-making ZWW.CW
 '- Film studios RDH.X
 '- Location lighting RLH.T
 '- Motion-picture cameras LOB.C
 '- Motion-picture projectors LOD.C
 '- Recording engineering LN
 '- Special effects (photography) LPO

Cineole DVO.P
 - Terpene hydrocarbons

Cinnamon IIG.I
 < Spices

Cinnamon IIG.I
 + Essential oils VMF/VMH
 =" Oil of cinnamon VMG.D

Circles (geometry) CCG.J
 - Arcs of a circle
 - Geometry

Circling guidance lights
 - Aeronautical ground lights RLH.S

Circlips NWU.FF
 < Spring retaining rings

Circuit analysis
 - Network analysis (circuits) KPE

Circuit-breaker components KJC.R
 > Arc control devices
 > Interrupters (circuit-breakers)
 > Tripping mechanisms (circuit-breakers)
 - Circuit-breakers

KJC → Circuit-breakers
= Air-break circuit-breakers
= Air circuit-breakers
* >Earth-leakage circuit-breakers KIP.Q
* >Relay circuit-breakers KIP.PC
* - Operating time MBC.DP
* - Switch-fuses KJH.C
* - Switches KJH

(By size)

KJC.C Miniature circuit-breakers
* – Fuses KIP.M

(By operating medium)

KJC.E Oil circuit-breakers
KJC.G Gas-blast circuit-breakers
KJC.GC Air-blast circuit-breakers
KJC.H Vacuum circuit-breakers
* <Vacuum devices NPT

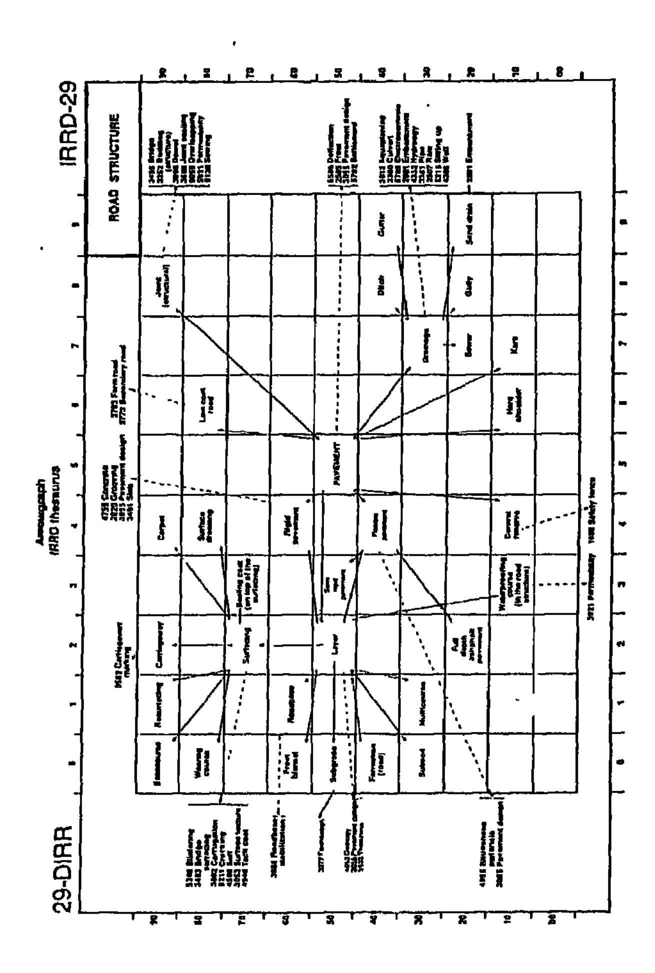
(By design)

KJC.M Tri-pole circuit-breakers
= Triple-pole circuit-breakers

→ Circuit-breakers KJC
= Air-break circuit-breakers
= Air circuit-breakers
< Switchgear
> Gas-blast circuit-breakers
> Miniature circuit-breakers
> Oil circuit-breakers
> Overcurrent circuit-breakers
> Tri-pole circuit-breakers
> Vacuum circuit-breakers
- Breaking capacity
- Circuit-breaker components
- Making capacity
- Recovery voltage
*> Earth-leakage circuit-breakers KIP.Q
*> Relay circuit-breakers KIP.PC
*- Operating time MBC.DP
*- Switch-fuses KJH.C
*- Switches KJH

Circuit design
- Network synthesis KPG

شكل (٥) عرض الأسهم



العرض فى المكانز متعددة اللغات:

يعتبر مكنز اليونسكو Unesco Thesaurus من أبرز المكانز متعددة اللغات. وقد نشر هذا المكنز متعدد اللغات فى مجلد واحد مؤلف من قسمين: القسم الهجائى وهو القسم الرئيسى والقسم المنهجى وهو بالإنجليزية ويمكن الوصول للمصطلحات باللغات الأخرى عن طريق الكشافات بالفرنسية والأسبانية بالمكنز.

ومن البدائل الأخرى: ظهور طبعة مستقلة لكل لغة حيث تظهر كل أشكال العروض الهجائية والمنهجية والتبادلية وغيرها بلغة الطبعة. وقد تشتمل الطبعة أيضاً على كشافات باللغات الأخرى للمكنز ومن الشائع أن يظهر العرض الهجائى الرئيسى فى المكنز متعدد اللغات المصطلحات المقابلة فى اللغات الأخرى تحت المصطلح المفضل مباشرة وقبل أية تبصرات توضيحية أو مصطلحات غير مفضلة [١٠].

رابعاً: خطوات إعداد المكنز وتحديثه [١١] :

(١) تحديد المجال الموضوعى :

إن نقطة البدء فى إعداد أى مكنز هى تحديد المجال الموضوعى الذى سيتم تغطيته، ويشمل ذلك: وضع حدود الموضوع، والتمييز بين النطاق المركزى Central Area الذى ينبغى أن يعامل بعمق والموضوعات الجانبية والتى قد تكون المكانز الأخرى المتوفرة كافية بالنسبة لها.

وبعد توضيح حدود الموضوع تراجع لغات التكشيف والاسترجاع المتاحة بالنسبة للموضوع الرئيسى أولاً ثم بالنسبة للموضوعات الجانبية بعد ذلك.

(٢) اختيار خصائص المكنز وشكله :

يجب على المسئول عن إعداد المكنز فى هذه المرحلة أن يكوّن رأيه فيما يتعلق بنوع المكنز المرغوب آخذاً فى الاعتبار احتياجات نظام المعلومات ككل، وأن يحدد

(10) Aitchison, Jean. Thesaurus construction and use, 1997. – p. 91-134.

(11) Atichison Jean. Thesaurus construction and use. 1997.- p.141-163.

Lancaster, F.W. Vocabulary control for information retrieval..- p 27-37.

هل يناسب المكنز نظام الربط المسبق أم الربط اللاحق أم كلاهما، وعلى المسئول أن يتخذ قراراته فيما يتعلق بأمور مثل: مستوى التخصيص، مستوى الربط المسبق، مدى العلاقات الهرمية وغيرها من العلاقات. كما أن على المسئول أن يقرر الشكل النهائى للمكنز، أى كيف سيرتب القسم الرئيسى وما مدى الحاجة للأقسام الأخرى.

(٣) اختيار المصطلحات :

يحسن قبل البدء فى اختيار المصطلحات تقسيم المجال الموضوعى إلى مجموعات أو أوجه رئيسية.

وهناك عدة طرق لجمع المصطلحات منها:

(أ) جمع المصطلحات تجريبياً على أساس تكشيف مجموعة ممثلة من الوثائق أو مصادر المعلومات.

(ب) تحويل أداة موجودة بالفعل، مثل تحويل قائمة رؤوس موضوعات إلى مكنز.

(ج) اقتباس مكنز من واحد آخر أكثر عمومية، أى إعداد مكنز مصغر.

(د) جمع المصطلحات من مصادر متعددة سواء من المعاجم وغيرها من المطبوعات أو من المتخصصين الموضوعيين.

وعادة ما تتبع الطريقة الثانية أو الطريقة الثالثة فى ظروف خاصة، بينما تعتبر الطريقة الأولى أو الطريقة الرابعة من الطرق شائعة الاستخدام بصفة عامة. وتسمى الطريقة الأولى بالطريقة التحليلية أو التجريبية، وهى تتضمن تحليل المحتوى الموضوعى للإنتاج الفكرى واختيار المصطلحات من الإنتاج الفكرى نفسه، وهى تفضل بصفة خاصة فى المجالات الموضوعية المتخصصة.

أما الطريقة الثانية فإن البعض يسميها طريقة اللجنة، حيث يتم تشغيل عدد من الخبراء يحصلون على المصطلحات من المصادر المختلفة، ويقومون باختيار المصطلحات المفضلة وإنشاء العلاقات المتبادلة بين المصطلحات وما إلى ذلك.

وتنطبق هذه الطريقة بصفة عامة على المجالات العريضة التى تتضمن موضوعات متعددة. ومع هذا، فغالباً ما يحدث نوع من الدمج أو الربط بين الطريقتين معاً عند التطبيق العملى. ونشير فيما يلى إلى المصادر التى يمكن الاعتماد عليها للحصول على المصطلحات:

- المصادر الأولية، مثل قوائم المصطلحات، خطط التصنيف، الموسوعات ومعاجم المصطلحات، الكشافات ونشرات المستخلصات، الكتب الدراسية .... إلخ.
- فحص الإنتاج الفكرى نفسه أو التكشيف الفعلى للوثائق.
- فحص أسئلة المستفيدين.
- معرفة الجامع بالموضوع وتآلفه مع المصطلحات.
- معرفة المستفيدين وخبراتهم الشخصية.

وتعتمد الطرق الآلية على اشتقاق قوائم بالكلمات آلياً من النصوص باستخدام أساليب مثل الربط الاحصائى Statistical Association حيث تختار المصطلحات التى تتردد فى الاستخدام أكثر من غيرها.

(٤) تسجيل المصطلحات :

يستخدم نموذج خاص على بطاقة أو فى شكل مقروء آلياً، وهو الأفضل، لتسجيل المصطلحات المختارة. وتشمل المعلومات المطلوبة بالنسبة لكل مصطلح:

- المصطلح.
- المرادفات، والأشكال الأخرى أو البديلة للمصطلح.
- المصطلحات الأعرض.
- المصطلحات الأضيق.
- المصطلحات المتصلة أو المرتبطة (غير الهرمية).
- المصدر (إذا أخذ المصطلح من قاموس، مكنز، إلخ).
- التبصرات التوضيحية والتعريفات (عند الضرورة).
- رقم تصنيف المجموعة الموضوعية العريضة إذا كانت الوسائل التصنيفية ستستخدم أثناء التجميع.

وتجدر الإشارة إلى أن المعلومات عن المصطلح وعلاقاته تضاف تدريجياً على النموذج أثناء عملية التجميع. انظر شكل (٦) بطاقة جمع المصطلحات.

المصطلح	رمز التصنيف
المرادفات وما فى حكمها	تعريفات
المصطلحات الأعرض	تبصرات توضيحية
المصطلحات الأضيق	
المصطلحات المرتبطة	المصدر

(شكل ٦ بطاقة أو نموذج لجمع مصطلحات المكنز)

(٥) البناء :

وتشمل هذه الخطوة :

(أ) التنظيم المبدئى للموضوعات التى يغطيها المكنز.

يجب تنظيم الموضوعات التى يغطيها المكنز سواء قبل أو أثناء فترة تجميع المصطلحات إلى فئات رئيسية حتى يمكن فرز المصطلحات المختارة وصفّها تحت الرؤوس الملائمة انتظاراً لتحليل أكبر. وهذه العملية مهمة لأنها ستكشف تغطية المكنز وتبين الفجوات الواجب تغطيتها وفى نفس الوقت فإن إحضار المصطلحات المتشابهة معاً سوف يسّهل تقرير علاقاتها البنائية.

(ب) تحليل وتجميع المصطلحات داخل الفئات العريضة.

وهنا يتم تحليل المصطلحات فى كل فئة عريضة مع النظر فى علاقتها بالمصطلحات الأخرى فى الفئة لإيجاد نمط بنائى.

وبعد ذلك يتم التحرير.

وقد يستخدم العرض الموضوعى فى الشكل المنهجى لإنتاج مكنز هجائى تقليدى

(٦) الترتيب والتحرير والمراجعة:

تعد الإحالات اللازمة وترتب حسب النظام المقرر. وهناك أيضاً عمليات التحرير والمراجعة وهى تشمل: مراجعة علاقات المصطلح بالمصطلحات الأخرى، مراجعة أرقام أو رموز المصطلحات، حذف المكررات، مراجعة الترتيب .... إلخ.

ويجب أن يكون جامعوا المكنز على اتصال بالخبراء الموضوعيين أثناء كل مراحل البناء ولا يجب أن ينهوا وينتجوا المكنز قبل موافقة هؤلاء الخبراء.

(٧) كتابة المقدمة :

من الضرورى كتابة مقدمة وافية تغطى كل أو معظم النقاط التالية:

- الغرض من المكنز.
- التغطية الموضوعية مع إشارة إلى المجالات الرئيسية والمجالات الهامشية.
- العدد الكلى لمصطلحات التكشيف المستخدمة مع بيان عدد المصطلحات المفضلة والمصطلحات غير المفضلة.
- ضبط المفردات: المعايير المستخدمة والقواعد المتبعة فيما يتعلق باختيار المصطلحات وصياغتها.
- البناء والعلاقات الداخلية: المعايير المستخدمة والقواعد المطبقة.
- طريقة عرض المصطلحات بالمكنز.
- قواعد الترتيب المستخدمة مع إشارة إلى المعايير المستخدمة فى حالة وجودها.
- معانى الاختصارات وعلامات الترقيم المستخدمة.
- الاستخدام العملى المكنز: كيف يستخدم المكنز فى البحث والتكشيف.
- التحديث والصيانة: تفاصيل سياسة التحديث واسم وعنوان الهيئة المسئولة عن صيانة المكنز.
- توجيه الشكر لأصحاب المصادر المستخدمة.
- توجيه الشكر للأفراد الذين ساعدوا فى تجميع المكنز.

ويجب توضيح هذه النقاط بالأمثلة اللازمة ما أمكن.

(٨) الاختبار :

لابد من إجراء عملى للمكنز قبل أن يصبح جاهزاً للعمل. ومن ثم يستخدم المكنز لتكشيف عدد مختار من الوثائق، كما يجب اختبار المكنز أمام الاستفسارات وطلبات البحث التى تقدم للنظام. وقد تكشف هذه الاختبارات عن الحاجة إلى إضافة مصطلحات جديدة أو إجراء بعض التعديلات.

(٩) الإعداد للنشر :

يجب عمل مسودة من المكنز للمناقشات والاختبار قبل النشر النهائى.

ونأتى أخيراً إلى الإنتاج النهائى للمكنز. وتعتمد الطريقة المختارة على عدد من العوامل مثل: عدد النسخ المطلوبة، وهل المكنز للاستخدام الداخلى المحدود أم للبيع والاستخدام على نطاق واسع.

ولا جدال فى أن استخدام الحاسوب فى إعداد المكنز سوف يريح من عمليات كتابية أو روتينية كثيرة، كما أنه يقلل من الأخطاء إلى حد كبير. وكل ما هنالك أن المصطلحات وما يرتبط بها من معلومات تسجل فى شكل مقروء آلياً بدلاً من البطاقات العادية فى الطريق اليدوية. وهناك البرامج التى تولّد من هذا السجل القوائم الهجائية والموضوعية، والإحالات والمداخل المتبادلة ... إلخ.

(١٠) تحديث المكنز :

يتطلب إعداد المكنز ونشره إجراءات سبق الإشارة إليها بايجاز، إلا أن الأمر لا يقف عند هذا الحد إذ يفقد المكنز حداثته بمجرد نشره، ولذلك فإن أى مكنز حى ينبغى أن يحافظ على حداثته بصفة مستمرة إذا كان له أن يستخدم كأداة فعالة فى التكشيف واسترجاع المعلومات. ويرجع السبب فى ذلك إلى نشأة مفاهيم ومصطلحات جديدة، أو اتضاح عدم استخدم بعض المصطلحات الموجودة بالفعل فى المكنز، وحتى إذا لم يحدث هذا فإن الفرد عند تكشيفه لعدد كبير من الوثائق يجد نفسه أمام مصطلحات كثيرة لم تلاحظ فى البناء الأول للمكنز.

ويجب إضافة واصفات جديدة إذا تبين أن هناك حاجة لها. فإذا قابل المكشف رأساً ليس فى المكنز ما يغطيه بكفاية فإنه يقدم اقتراحاً بضم واصف جديد. ويعتمد التبرير للواصف الجديد على خبرة المكشف من ناحية وقواعد المكنز من ناحية ثانية، بالإضافة إلى القواميس والمواد المرجعية الأخرى. ويوضع الواصف الجديد فى نموذج معد لهذا الغرض ومعه إحالاته التى تبين ارتباطاته بالمصطلحات الموجودة بالمكنز كذلك أيضاً تبصره توضيحية إذا لزم الأمر. وتعرض نماذج الواصفات الجديدة على الأخصائيين الموضوعيين واللغويين لدراستها على ضوء

قواعد المكنز فإن تمت الموافقة على الواصف فإنه يصبح جاهزاً للإدخال فى المكنز. وقد يظهر فى نشرة خاصة بالإضافات والتعديلات توزع على المستفيدين من المكنز.

وبجانب إضافة مصطلحات جديدة فإن تحديث المكنز يشمل أيضاً حذف مصطلحات موجودة وإن كانت كمية الحذف لا تساوى كمية الإضافة. فإذا وجد أن أحد المصطلحات لا يستخدم إلا نادراً فإنه ينبغى التحقق مما إذا كان ذلك راجعاً إلى نقص الوثائق المتعلقة بهذا المصطلح أو أنه مجرد تغير فى التسمية. وهكذا فقد يحذف المصطلح ويحل محله مرادف أو يحذف ويحال منه بإحالة انظر إلى مصطلح آخر ذى نطاق أعرض. ولا ينبغى أن يستبعد كلية إلا إذا كان لم يستخدم على الإطلاق.

وقد يتطلب الأمر أيضاً تصحيح مصطلح موجود، وحذف أو إضافة علاقات للمصطلحات وما إلى ذلك. وجدير بالذكر أن إضافة أحد المصطلحات أو حذف مصطلح أو تعديل مصطلح ليست عملية قاصرة على المصطلح نفسه فحسب وإنما ينبغى أن تمتد إلى كل المصطلحات الأخرى ذات الصلة به التى يشتمل عليها المكنز

وعلى أى الأحوال فإنه يجب عمل مراجعة دورية للتحقق من مدى نفعية أو قيمة الواصفات سواء لأغراض التكشيف أو الاسترجاع، ويجب أن تؤكد المراجعة الدورية أن الواصفات لا تتعارض مع بعضها البعض أو يكرر أحدها الآخر وأن العلاقات المنشأة بين المصطلحات فى المكنز ما تزال صحيحة ويمكن استخدامها (١٢).

وقد يبدو من الضرورى إجراء مراجعة شاملة للمكنز بعد مضى عدد من السنوات عليه بسبب التغييرات الكثيرة التى أجريت فيه من وقت لآخر. ويلعب الحاسب الالكترونى دوراً كبيراً فى عملية تحديث المكنز وصيانته وهو يجعلها أكثر دقة وسهولة وسرعة.

---

(12) Aitchison, Jean. Thesaurus construction... 1997.-p. 166.

Unesco. Guidelines for establishment and development of monlingual thesauri... p. 34-36.

خامساً : الإسهامات العربية فى مجال المكانز :

يمكن التعرف على المساهمات العربية فى مجال المكانز من خلال دليل الإنتاج الفكرى العربى فى مجال المكتبات المعلومات بمجلداته المتعددة والتى تغطى الإنتاج منذ أواخر القرن التاسع عشر حتى أواخر القرن العشرين [١٣].

وقد جرى فحص هذا الدليل فتبين وجود ١٠٩ مادة عن المكانز حتى عام ١٩٩٨. وقد تم استبعاد ٤٥ مادة منها بعد أن تبين أن معظمها مواد إخبارية أو تقريرية نشرت فى مجلات إخبارية وخاصة الرسالة الإخبارية للشبكة العربية للمعلومات (تغير اسمها فيما بعد إلى الرسالة الإخبارية لمركز التوثيق والمعلومات بجامعة الدول العربية)

١/٥ بعض السمات البيليوجرافية للإنتاج الفكرى العربى عن المكانز:

يمتد عمر الإنتاج الفكرى على مدار نحو عشرين عاماً ابتداء من عام ١٩٧٨ حين نشرت أول دراسة عن الموضوع وهى مقالة عن نشأة المكانز كأدوات للتكشيف واسترجاع المعلومات لمحمد فتحى عبد الهادى [١٤]. وقد شهدت السبعينات صدور خمس مواد بينما شهدت الثمانينات صدور ٣٠ مادة والتسعينات ٢٩ مادة.

---

(١٣) انظر بصفة خاصة: محمد فتحى عبد الهادى. الإنتاج الفكرى العربى فى مجال المكتبات والمعلومات فى عشر سنوات، ١٩٧٦ ــ ١٩٨٥. – الرياض: دار المريخ للنشر، ١٩٨٩.

محمد فتحى عبد الهادى. الإنتاج الفكرى العربى فى مجال المكتبات والمعلومات، ١٩٨٦–١٩٩٠م. ــ الرياض: مكتبة الملك فهد الوطنية، ١٩٩٥.

محمد فتحى عبد الهادى. الإنتاج الفكرى العربى فى مجال المكتبات والمعلومات، ١٩٩١–١٩٩٦م.– الرياض: مكتبة الملك فهد الوطنية، ٢٠٠٠.

(١٤) محمد فتحى عبد الهادى. المكنز كأدوات للتكشيف واسترجاع المعلومات: ١– النشأة. – صحيفة المكتبة. – مج ١٠، ع ٢ (إبريل ١٩٧٨). ــ ص ٥–١٢.

وتتوزع فئات المواد على النحو التالى (انظر جدول ١)

الفئة	العدد
كتب	٢١
مقالات ودوريات	٣٣
رسائل جامعية	٥
دراسات مقدمة إلى مؤتمرات	٥
المجموع	**٦٤**

جدول (١) فئات المواد عن المكانز

وقد يبدو أن عدد الكتب كبيراً لكن حقيقة الأمر أن هذا العدد يضم فئتين متميزتين هما المكانز (١٧) والكتب الإرشادية (٤)، أما المقالات فهى موزعة على عدد كبير من الدوريات العربية، إذ بلغ عدد الدوريات التى نشرت مقالات عن المكانز ١٥ دورية أبرزها: المجلة العربية للمعلومات (تونس) ١١ مادة ثم مجلة المكتبات والمعلومات العربية (٣ مواد) ومجلة رسالة المكتبة (٣ مواد).

ويلاحظ قلة عدد الرسائل الجامعية التى تناقش قضايا المكانز فى المكتبات ومراكز المعلومات العربية، إذ تم رصد خمس رسائل، اثنتان للدكتوراه وثلاثة للماجستير، أقدمها رسالة ماجستير قدمت عام ١٩٨٣ من باحثة مصرية (أمنية مصطفى صادق) إلى جامعة شفيلد ببريطانيا وهى تتناول بالتقييم والدراسة ستة مكانز بالإنجليزية فى مجال العلوم الاجتماعية. كما أن أول رسالة عربية عن إعداد المكانز العربية هى رسالة دكتوراه قدمها شكرى العنانى عام ١٩٩١.

ورغم أن الإنتاج الفكرى عن المكانز يتوزع جغرافيا على ١١ دولة إلا أن أبرز الدول التى قدمت مساهمات عربية عن المكانز هى تونس والسعودية ومصر إذ قدمت هذه الدول (٤١) مادة من مجمل (٦٤) ويشير استعراض أسماء المؤلفين إلى بروز ثلاثة مؤلفين فقط فى هذا المجال هم محمد فتحى عبد الهادى ومحمود وأتيم وشوقى سالم. فقد شاركوا فى إعداد مكانز عربية، كما قدموا العديد من المقالات والدراسات حول الموضوع.

Converted by Tiff Combine - (no stamps are applied by registered version)

٢/٥ كلمة مكنز العربية :

إن كلمة Thesaurus الإنجليزية قد ترجمت إلى العربية بعدة أشكال منها:

قائمة مصطلحات (١٩٧٠)، قائمة أبجدية للألفاظ (١٩٧٢)، مكنز (١٩٧٥) كما عربت الكلمة بـــ الثيسورس (١٩٨٢).

ومن المؤكد أن كلمة "مكنز" كمقابل للكلمة الإنجليزية Thesaurus هى أفضل المقابلات وأحسنها، وقد استخدمها سعد الهجرسى فى دراسة له عن استخدام الحاسب الإلكترونى فى الأعمال الببليوجرافية بالوطن العربى عام ١٩٧٥، وأول دراسة عن الموضوع تحتوى فى عنوانها على كلمة مكانز هى المقالة التى نشرها محمد فتحى عبد الهادى عام ١٩٧٨ بعنوان المكانز كأدوات للتكشيف واسترجاع المعلومات. وليس من المستحسن استخدام اللفظ المعرب الثيسورس كمقابل رغم وروده فى مقال لصالح الزيناتى نشره فى رسالة المكتبة عام ١٩٨٢. فشكل اللفظ غير مقبول، فضلاً عن أن كلمة "المكنز" قد شاع استخدامها فى الكتابات المتخصصة عن الموضوع منذ أواخر السبعينات من القرن العشرين.

٣/٥ المواصفات القياسية العربية للمكانز:

أعدت المنظمة الدولية للتقييس (ISO) مواصفتين للمكانز إحداهما لإعداد وتطوير المكانز أحادية اللغة والثانية لإعداد وتطوير المكانز متعددة اللغات.

وقد تم ترجمة أو تعريب هاتين المواصفتين إلى العربية عن طريق المنظمة العربية للمواصفات والمقاييس قبل إلغائها وصدرت الأولى عام ١٩٨٤ والثانية عام ١٩٨٦. ورغبة فى توسيع الإفادة من هاتين المواصفتين على نطاق واسع قامت المنظمة العربية للتربية والثقافة والعلوم بنشرها فى مجلتها المعروفة "المجلة العربية للمعلومات"، المواصفة الأولى عام ١٩٨٥ والثانية عام ١٩٨٧ ثم أعادت نشرهما مرى أخرى فى العدد الثانى من المجلد الثالث عشر الصادر عام ١٩٩٢ من المجلة "استجابة لطلبات عديدة من قبل طلبة أقسام علم المكتبات والمعلومات والعاملين فى المكتبات ومراكز التوثيق العربية".

وتشتمل النصوص العربية على أمثلة ونماذج عربية إضافة إلى الأمثلة والنماذج باللغات الأخرى.

ومن الملاحظ أن هذه المواصفات لم تستخدم على نطاق واسع سواء فى إعداد المكانز العربية أو فى التدريس بأقسام المكتبات والمعلومات، كما يلاحظ أنه لم يتم تحديثها رغم صدور طبعة جديدة من المواصفة الدولية للمكانز أحادية اللغة عام ١٩٨٦.

٤/٥ أربعة كتب عربية عن المكانز :

عقد مركز التوثيق والمعلومات بالأمانة العامة لجامعة الدول العربية دورة تدريبية للموثقين العرب حول بناء المكانز وتطويرها فى دمشق خلال الفترة من ٣-١٥/٨/١٩٨٥ شارك فى إلقاء المحاضرات فيها كل من ف. ر. لانكستر ومحمود أحمد أتيم واعتماداً على ما ألقى من محاضرات فى هذه الدورة أعد كتاب "بناء المكانز وتطويرها" وهو من تحرير محمود أتيم وصدر فى تونس عام ١٩٨٧ [١٥] .

ويقوم هذا الكتاب على محاضرتين أولاهما عن بناء المكانز واستخدامها وهى التى أعدها العالم المشهور فردريك لانكستر، والثانية عن المكانز متعددة اللغات وخصوصيات اللغة العربية وهى التى أعدها محمود اتيم. ثم هناك مجموعة من الملاحق أهمها المواصفتان المعربتان عن إعداد المكانز، ومواصفة معربة عن طرق تفحص الوثائق وتعيين موضوعاتها واختيار مصطلحات التكشيف.

وفى عام ١٩٨٩ صدر أول كتاب متكامل عن الموضوع بعنوان "المكانز كأدوات للتكشيف واسترجاع المعلومات" من إعداد محمد فتحى عبد الهادى [١٦]. ويشتمل الكتاب فى معظمه على مجموعة من الدراسات التى سبق أن نشرها المؤلف فى عدد من الدوريات العربية المتخصصة مع تعديلات وإضافات. ويتناول الكتاب المكانز من كافة النواحى سواء من حيث نشأتها وتعريفها ووظائفها وأنواعها أو من حيث بناء المكنز كما يتمثل فى تقنين وضبط المصطلحات أو العلاقات بين المصطلحات أو تنظيم وعرض المصطلحات فى المكانز كما يتناول الكتاب خطوات إعداد المكانز وطرق تحديثها وصيانتها. وإضافة إلى ذلك يوجد فصل يشتمل على

---

(١٥) جامعة الدول العربية. الأمانة العامة. مركز التوثيق والمعلومات. بناء المكانز وتطويرها/ إشراف محمود أحمد أتيم. – تونس: المركز، ١٩٨٧.- ٣٧٢ص.

(١٦) محمد فتحى عبد الهادى. المكانز كأدوات للتكشيف واسترجاع المعلومات. – القاهرة: مكتبة غريب، ١٩٨٩. – ١٦٧ص

معلومات وصفية عن أهم المكانز وفصل آخر يتضمن إرشادات ومعايير إعداد المكانز وتقييمها.

والكتاب الثالث هو ترجمة عربية للطبعة الثانية من كتاب لانكستر القيم بعنوان "ضبط المفردات من أجل استرجاع المعلومات" لمحمود أتيم وقد صدر الكتاب المترجم بعمان عام ١٩٩٣ (١٧). ولهذا الكتاب أهميته وقيمته الكبيرة فهو من إعداد علم بارز فى علم المعلومات بصفة عامة والمكانز بصفة خاصة. ويتميز الكتاب بحداثة المعلومات وشمولية التغطية، فضلاً عن اشتماله على عديد من الأشكال والنماذج التوضيحية.

أما الكتاب الرابع فهو كتاب هانى محيى الدين عطية الذى صدر بالقاهرة عام ١٩٩٧ بعنوان "نحو منهج لتنظيم المصطلح الشرعى: مدخل معرفى معلوماتى" (١٨). ورغم أن العنوان يبدو غريباً إلا أن الكتاب من أوله إلى آخره يتناول موضوع المكانز. ويتمثل عمل المؤلف فى دراسة خصائص اللغة الإصطلاحية الشرعية ويرتكز فيه بعد ذلك على تطبيق مفهوم المكنز كأداة استرجاع معلومات على تلك المصطلحات. ويضم الكتاب ستة أبواب تتناول المصطلح والمكنز وضبط وتقنين المصطلحات فى المكنز والعلاقات بين المصطلحات فى المكنز وتنظيم وعرض المصطلحات فى المكنز وبناء وتطوير المكانز، كما تم استعراض الجهود المعاصرة فى إعداد المكانز الإسلامية.

ويتميز الكتاب باهتمام مؤلفه بدراسة المصطلحات فى اللغة العربية واعتماده على عدد من المصادر العربية التراثية.

٥/٥ نشأة المكانز العربية وتطورها:

يبدو أن أول عمل صدر بالعربية هو قائمة المصطلحات العربية الصادرة عن مركز التنمية الصناعية للدول العربية (المنظمة العربية للتنمية الصناعية حالياً)

---

(١٧) لانكستر، فريدريك ويلفريد. ضبط المفردات من أجل استرجاع المعلومات، الطبعة الثانية/، ترجمة محمود أحمد أتيم. – عمان: جمعية المكتبات الأردنية، ١٩٩٣. – ٢١٤ص

(١٨) هانئ محيى الدين عطية. نحو منهج لتنظيم المصطلح الشرعى: مدخل معرفى معلوماتى.– القاهرة: المعهد العالمى للفكر الإسلامى، ١٩٩٧. – ٢٦٠ ص

بالقاهرة عام ١٩٧٤، وهذه القائمة تم تطويرها فى عام ١٩٧٩ وصدرت بعنوان المكنز الشامل للمصطلحات فى مجالات التنمية الاقتصادية والاجتماعية [١٩] وهو فى الأساس تعريب لمكنز Macrothesaurus الصادر عن منظمة التعاون الاقتصادى والتنمية التابعة للأمم المتحدة، وتتابع بعد ذلك صدور المكانز العربية أحادية اللغة أو متعددة اللغات. وعلى حد علم الباحث فقد وصل عدد المكانز العربية إلى نحو عشرين مكنزاً حتى أواخر التسعينات من القرن العشرين.

وقد قسمت سوزان فلمبان فى رسالتها للماجستير [٢٠] المراحل التى مر بها إعداد المكانز العربية على النحو التالى:

١- المرحلة الأولية

وهى التى تنسب إليها ظهور المكانز العربية فى السبعينات ابتداء من المكنز الشامل للمصطلحات فى مجالات التنمية الاقتصادية والاجتماعية (١٩٧٤، ١٩٧٩)

٢- مرحلة التوصيات والمنهجية :

وقد تميزت هذه المرحلة (النصف الأول من الثمانيات) بصدور الأسس المنهجية لإعداد المكانز، كما شهدت صدور بعض المكانز أولها مكنز مصطلحات علم المكتبات والمعلومات لمحمد فتحى عبد الهادى عام ١٩٨٠.

٣- مرحلة النشاط :

تميزت هذه المرحلة (النصف الثانى من الثمانيات والنصف الأول من التسعينات) بكثرة إنتاج المكانز العربية وذلك بعد ثبات واستقرار مفهوم استخدام المكانز فى التحليل الموضوعى الدقيق والاسترجاع. ومن أهم نماذج هذه المرحلة مكنز الجامعة الصادر عن جامعة الدول العربية عام ١٩٨٧ ومكنز الفيصل الصادر عن مركز الملك فيصل للبحوث والدراسات الإسلامية عام ١٩٩٤.

---

(١٩) المكنز الشامل للمصطلحات فى مجالات التنمية الاقتصادية والاجتماعية: النسخة العربية. – القاهرة: مركز التنمية الصناعية للدول العربية، ١٩٧٩.

(٢٠) سوزان مصطفى فلمبان. المكانز كأدوات للتحليل الموضوعى: دراسة تقويمية لاثنين من المكانز العربية. ــ جدة، ١٩٩٥، ص ٤٣-٥٤. (أطروحه ماجستير ــ جامعة الملك عبد العزيز. كلية الآداب، قسم المكتبات والمعلومات)

٤- مرحلة البرامج الجاهزة لإعداد المكانز العربية :

تميزت هذه المرحلة بالاستفادة الكبيرة من إمكانيات الحاسب الالكترونى واستغلالها فى إنجاز أغلب عمليات بناء وتطوير المكانز بشكل يوفر الوقت ويقلل من الجهد البشرى ويضمن الدقة والكفاءة فى تكوين المكنز وحيث اهتمت بعض المؤسسات العربية بتطوير نظم آلية جاهزة لإعداد المكانز العربية تقوم فكرتها على أساس ميكانيكية تغذية المكنز بالمصطلحات، وما تزال هذه المرحلة التى بدأت فى التسعينات فى بدايتها.

ونسجل فيما يلى بعض الملاحظات العامة على المكانز العربية:

(أ) رغم عدم وجود حصر دقيق بالمكانز الصادرة بالعربية إلا أن العدد (حوالى ٢٠) قليل بصفة عامة إذا أخذنا فى الاعتبار ما صدر من المكانز فى اللغات الأخرى وخاصة الإنجليزية وإذا أخذنا فى الاعتبار كذلك المجالات الموضوعية العديدة التى تحتاج إلى مكانز متخصصة.

(ب) تغطى المكانز الصادرة بالعربية المجالات الموضوعية التالية :

التنمية الاقتصادية والاجتماعية

التربية

المكتبات والمعلومات

العلوم الإسلامية

السكان

العمل

البترول

الطفولة

الزراعة

وهكذا فالمجالات محدودة وتنتمى معظم المكانز إلى قطاع العلوم الاجتماعية والإنسانيات، أما المكانز العامة أو شبه العامة فهى ثلاث:

مكنز الجامعة ومكنز اليونسكو والمكنز الموسع.

(ج) تعتمد معظم المكانز العربية على الترجمة من مكانز بلغات أخرى غير العربية وخاصة الإنجليزية وبعضها يدرج المقابلات الأجنبية والبعض الآخر لا يدرجها. ومن أبرز المكانز المترجمة:

مكنز اليونسكو

أما المكانز عربية النشأة فهى محدودة للغاية وأبرزها مكنز الفيصل. وسوف نتناول مسألة التاليف والترجمة فى نقطة تالية.

(د) رغم صدور نحو عشرين مكنزاً إلا أن استخدامات هذه المكانز محدودة، وهى تتقادم من وقت لآخر وتعانى من عدم التحديث وهى فى شكل مطبوع ولم يتح معظمها بعد فى شكل محسَب أو مليزر، وعادة ما يذكر فى مقدمات المكانز أن الطبعة تجريبية أو تمهيدية وأنه لابد من متابعتها أو تحديثها بصورة مستمرة إلا أن ذلك لا يحدث، إلا فى أحوال قليلة.

٥/٦ المكانز العربية بين التأليف والترجمة :

سبق أن أشرنا إلى أن معظم المكانز العربية مترجمة عن أصول أجنبية ومن أبرزها مكنز "الجامعة" الصادر عن مركز التوثيق والمعلومات بجامعة الدول العربية [٢١].

وهو مكنز يضم الموضوعات التى يغطيها نشاط جامعة الدول العربية ويعكس برامجها. وتتمثل أهمها فى الموضوعات ذات الصلة بالشؤون السياسية والعلاقات الدولية والتسلح ونزع السلاح والتجارة الدولية والتنمية الاقتصادية والاجتماعية والزراعية والتربية والتعليم والعمل والعمال والسكان والصحة والإسكان والصناعة والعلوم والتكنولوجيا.

ولعله يتضح مما سبق مدى اتساع حدود التغطية الموضوعية ولذلك فقد وصل عدد المصطلحات بالمكنز إلى أكثر من عشرة آلاف مصطلح (١٠٠٦٠ مصطلح).

(٢١) جامعة الدول العربية. الأمانة العامة. مركز التوثيق والمعلومات. الجامعة مكنز ثلاثى اللغات: العربية، الإنجليزية، الفرنسية. - ط١ (العربية). - تونس: المركز، ١٩٨٧ - ٢ مج .

وقد بنى مكنز "الجامعة"، الطبعة العربية، أساسا على مكنز UNBIS (United Nations Bibliographic Information System) الذى أعد فى مكتبة داغ همرشلد فى الأمم المتحدة.

وعند تعريب المكنز السابق لتكشيف الوثائق باللغة العربية تم فى البداية حذف المصطلحات التى لا تدخل فى نطاق اهتمامات الجامعة، وشرُع بعد ذلك فى عملية التعريب التى تمت اعتماداً على الواصفات المرتبة حسب الفئات الموضوعية واستخدام المعاجم المتخصصة فى ترجمتها، وتم استخراج كافة المترادفات الموجودة فى المعاجم ثم عرضت الترجمة على أخصائيين موضوعيين لاعتماد الواصفات من بين المترادفات واعتبار المترادفات الأخرى لاواصفات تم تمثيلها بشكل إحالات عربية أى بدون مقابل انجليزى أو فرنسى لها. وبهذه الطريقة أضيفت حوالى ٢٠٠٠ لاواصفة عربية كانت بعضها مترادفات لغوية وبعضها يخضع للاستخدام القطرى أو الاقليمى.

(شكل ٧ نماذج من مكنز الجامعة)

1 ــ القائمة الهجائية

ابحاث التغذية .... الواصفة العربية
NUTRITION RESEARCH .... الواصفة الانجليزية
RECHERCHE NUTRITIONNELLE .... الواصفة الفرنسية
12.01.00 .... رقم الوجه
س ل : بحوث التغذية .... استخدم ل
م ع : البحث العلمي .... مصطلح اعم
م ض : ابحاث البروتين .... مصطلح اضيق
م ت : ابحاث الخدمات الصحية .... مصطلح مترابط
الابحاث الطبية
الابحاث الغذائية

* بحوث التغذية .... اللاواصفة
12.01.00 .... رقم الوجه
س : ابحاث التغذية .... إستخدم

2 ــ كشاف الكلمات المفتاحية خارج السياق

ابحاث .... الكلمة المفتاحية
09.01.02 ابحاث الارز
16.03.02 ابحاث الارصاد الجوية
07.01.03 ابحاث الاسواق
16.03.01 الابحاث الاقيانوغرافية
14.07.01 ابحاث الإتصال
10.01.00 الابحاث الإجتماعية
04.04.02 الابحاث الإدارية
15.05.00 الابحاث التربوية
16.03.01 الابحاث التعدينية
12.01.00 ابحاث التغذية .... الواصفة

تغذية .... الكلمة المفتاحية
12.01.00 أبحاث التغذية .... الواصفة
12.02.00 أمراض التغذية
12.01.00 إحصاءات التغذية
12.01.00 برامج التغذية
12.01.00 تعليم التغذية
12.01.00 التغذيــة
09.03.00 تغذية الحيوان
12.01.00 تغذية الرضع
12.01.00 تغذية الطفل
09.02.01 تغذية النباتا

3 ــ القائمة المصنفة

12 07.00 .... رقم الوجه

أبحاث التغذية .... الواصفة

الأبحاث الغذائية

أسعار الأغذية

الأغذيـة

الأغذية الغنية بالبروتين

الأغذية القابلة للتلف

• الأغذية وتأثير الإشعاع عليها

س : تلوث الأغذية

• بحوث التغذية .... اللاواصفة

س : أبحاث التغذية

4 ــ كشاف إنجليزي ــ فرنسي ــ عربي

التغذيـة	NUTRITION	NUTRITION
تعليم التغذية	EDUCATION NUTRITIONNELLE	NUTRITION EDUCATION
سياسة التغذية	POLITIQUE NUTRITIONNELLE	NUTRITION POLICY
برامج التغذية	PROGRAMMES DE NUTRITION	NUTRITION PROGRAMMES
أبحاث التغذية	RECHERCHE NUTRITIONNELLE	NUTRITION RESEARCH
إحصاءات التغذية	STATISTIQUES NUTRITIONNELLES	NUTRITION STATISTICS
أمراض التغذية	MALADIES DE LA NUTRITION	NUTRITIONAL DISEASES

5 ــ كشاف فرنسي ــ إنجليزي ــ عربي

الأبحاث الطبية	MEDICAL RESEARCH	RECHERCHE MEDICALE
أبحاث الأرصاد الجوية	METEOROLOGICAL RESEARCH	RECHERCHE METEOROLOGIQUE
الأبحاث العسكرية	MILITARY RESEARCH	RECHERCHE MILITAIRE
الأبحاث التعدينية	MINING RESEARCH	RECHERCHE MINIERE
الأبحاث النووية	NUCLEAR RESEARCH	RECHERCHE NUCLEAIRE
أبحاث التغذية	NUTRITION RESEARCH	RECHERCHE NUTRITIONNELLE
الأبحاث الأقيانوغرافية	OCEANOGRAPHIC RESEARCH	RECHERCHE OCEANOGRAPHIQUE

وفيما يتصل بتعديل الفئات الموضوعية أو الأوجه فقد تبين أن القسم المصنف من مكنز UNBIS غير مناسب لاهتمامات الأمانة العامة لجامعة الدول العربية ولذلك تم إعادة هيكلة الأوجه وتبعا لذلك تم تقسيم بعضها وكذلك دمج بعضها الآخر حتى تغطى كل موضوعات اهتمامات إدارات الأمانة العامة.

وينقسم مكنز "الجامعة" إلى أربعة أقسام رئيسية (انظر شكل ٧):

١- قائمة هجائية مرتبة حسب الواصفات العربية.

٢- قائمة مصنفة.

٣- كشاف عربى بالكلمات المفتاحية خارج السياق.

٤- كشافات باللغات الثلاث.

ويعرض القسم الهجائى كل مصطلح فى هيكله الكامل فتشتمل الواصفات على العلاقات التكافئية والهرمية والترابطية مع التبصرات ورقم الوجه. وتشتمل اللاواصفات على العلاقة التكافئية ورقم الوجه. أما كشاف الكلمات المفتاحية خارج السياق فهو عبارة عن قائمة هجائية بالكلمات التى تتكون منها الواصفات، فنجد تحت كل كلمة الواصفات التى تكون هذه الكلمة المفتاحية جزءاً منها.

ويحدد القسم المصنف الفئات الفرعية ويعرض جميع المصطلحات التى تدخل فى نطاق هذه الفئة الفرعية، ولا يتضمن هذا القسم العلاقات بين المصطلحات ما عدا التبصرات وإحالات استخدم.

أما بالنسبة للكشافات باللغات الثلاث فقد ورد كشافان:

الأول مرتب باللغة الإنجليزية مع المقابلات بالفرنسية والعربية، والثانى مرتب باللغة الفرنسية مع المقابلات بالإنجليزية والعربية.

وقد ثارت مناقشات كثيرة حول إعداد المكانز العربية وهل الترجمة هى الأفضل أم أنه من الأحسن إنشاء المكانز إنشاءً عربياً أصيلاً، ومن حبذ الترجمة رأى أن الترجمة هى فى العادة لأصل تم إعداده بعناية ودراية كبيره ويستند إلى سنوات طويلة من العمل والخبرة، كما أن الترجمة تسهل إدخال اللغة العربية ضمن اللغات الأخرى فى المكانز متعددة اللغات، وهى مطلوبة فى البلاد العربية. ومن يعارض الترجمة يرى أن من عيوبها [٢٢] أن المكانز الأجنبية تعكس تحيزاً أو جهلاً فيما يخص الواقع العربى أو الإسلامى وخاصة فى قطاعات مثل الدين والإنسانيات، وأن المكانز الأجنبية قد وضعت لأغراض خاصة بها سواء بالنسبة للموضوعات التى تشتمل عليها أومسمياتها أو حاجات المستفيدين، هذا فضلاً عن أن هناك العديد من المسائل اللغوية الخاصة باللغة العربية منها: الاستخدامات القطرية المتعددة، وتفاوت عدد المقابلات بين اللغة العربية واللغات الأجنبية.

ولا شك أنه من الأفضل بالطبع إنشاء المكانز العربية وفقاً لخصوصية اللغة العربية مع الأخذ فى الاعتبار لطبيعة الموضوعات وأوجهها العربية من ناحية واحتياجات المستفيدين فى المكتبات ومراكز المعلومات العربية من ناحية أخرى. إلا أن هذا لا يمنع من الاستفادة من المكانز الأجنبية.

فإذا انتقلنا إلى المكانز المعدة بالعربية أصلاً فإن النماذج القديمة منها كانت خليطاً من الترجمة والتأليف، كما أن البعض الآخر كان مجرد نماذج مصغرة يأمل أصحابها فى الاقتداء بها عند إعداد المكانز العربية.

ومن ذلك ما قام به حسين الهبائلى (١٩٨٨) [٢٣] بإعداد مكنز مصغر نموذج فى مجال علم المعلومات ثلاثى اللغة (عربى ــ فرنسى ــ إنجليزى) واعتمد فى عرضه

(٢٢) محمود أحمد أتيم. إعداد المكانز وتطويرها. – المجلة العربية للمعلومات. – مج ٥، ع٢ (١٩٨٤). – ص ١٠٢.

(٢٣) حسين الهبائلى. المكانز متعددة اللغات من النظرية إلى التطبيق: الواقع والطموح. – المجلة العربية للمعلومات. – مج ٩، ع٢ (١٩٨٨).- ص ١٠٠- ١٢٩.

لهذا المكنز طريقة الأشكال السهمية التى يرى أنها تتلاءم تلاؤماً تاماً والمكانز متعددة اللغات.

وقد اشتملت هذه القائمة المتخصصة فى علم المعلومات على ١٥٨ مصطلحاً وبعد ترتيب القائمة هجائياً (وفقاً للمصطلح الإنجليزى) تبين أن هذه المجموعة تشكل مكنز مصغراً متخصص فى علم المعلومات يتكون من أربعة حقول دلالية هى: علم المعلومات، أوعية المعلومات، المعالجة الوثائقية، الإعلامية وكذلك من أربع قوائم إضافية هى:

المؤسسات والهيئات، المصطلحات الجغرافية، المجالات، الكلمات الوسيطة. ثم قام المؤلف بتوزيع المصطلحات على الحقول الدلالية وترتيبها هجائياً تحت كل منها بالعربية أولاً ثم المقابلات الفرنسية والإنجليزية، وبعد ذلك قام بهيكلة المصطلحات بطريقة الأشكال السهمية وربط العلاقات الدلالية بين الأشكال السهمية، ثم قام بإعداد القوائم الإضافية الأربع وهى قائمة المؤسسات والهيئات، قائمة المصطلحات الجغرافية، قائمة المجالات، قائمة الكلمات الوسيطة.

وهكذا تتناول عمل حسين الهبائلى عرض نموذج مصغر متعدد اللغات (عربى – فرنسى – إنجليزى) فى مجال علم المعلومات بطريقة الأشكال السهمية ليكون نبراساً فى طريقة إعداد المكانز العربية والمكانز متعددة اللغات ذات المدخل العربى. وفى رسالته للدكتوراه عن إعداد المكانز العربية، وهى أول رسالة عربية عن الموضوع، قام شكرى العنانى فى عام ١٩٩١ [٢٤] بإعداد مكنز عربى عن الراديو والتليفزيون بهدف أن يكون نموذجاً يمكن الاقتداء به فى إعداد مكنز عربى شامل يغطى مجال الاتصال الجماهيرى ككل. وقد بلغ عدد الواصفات فيه ٧٤٩ واصفة كما بلغ عدد اللاواصفات ٢٥٨ لاواصفه، وقد شرح العنانى فى مقدمة المكنز مراحل إنشائه، وأشار إلى أنه اعتمد فى بنائه على نظام مطور للحاسب الآلى يدعى المستشار صممته وطورته شركة النظم العربية المتطورة بالرياض. وقد تم التطوير لهذا النظام باستخدام قاعدة المعلومات العالمية أوراكل Oracle والتى يمكن تشغيلها

(٢٤) شكرى عبد السلام العنانى. إعداد المكانز العربية: دراسة تطبيقية فى مجال الاتصال الجماهيرى. – القاهرة، ١٩٩١. (أطروحة دكتوراه – جامعة القاهرة. كلية الآداب. قسم المكتبات والوثائق).

على غالبية أجهزة الحاسب الآلى المستخدمة فى المنطقة العربية. حيث قام الباحث إدخال كافة المصطلحات الأساسية التى تم جمعها بعلاقاتها الأساسية (الهرمية والترابطة)، وفور إدخال أى بطاقة لمصطلح جديد يقوم النظام بإضافة هذه العلاقات تحت المصطلح المقصود، وفى حالة خلو قاعدة المكنز من هذا المصطلح فإنه يقوم بفتح بطاقة جديدة له ويضيف إليها العلاقة المذكورة، وعقب انتهاء الباحث من تخزين كامل البطاقات تم طبع مخرجات للمكنز بصورة مبدئية لاستيفاء الإضافات اللازمة لبعض علاقات المصطلحات ومراجعة أخطاء الطباعة وصولاً بهذه المخرجات إلى الشكل النهائى. ويحتوى المكنز على عدد من العروض هى: العرض الهجائى، والعرض الهرمى ثم نموذج كشاف الكلمات المفتاحية خارج السياق، حيث تم اختيار حرف الألف من المكنز الذى بلغت واصفاته ١٦٩ واصفة ليكون الجزء الذى يتم به عمل نموذج كشاف الكلمات المفتاحية خارج السياق. وقد بلغ عدد الكلمات المفتاحية التى وردت بالنموذج ٣٦٩ كلمة وقد أظهرت التجربة عدداً من الملاحظات الإيجابية والسلبية.

وفى نفس هذا الاتجاه أيضاً قدم على خليفة فى رسالته للماجستير عام ١٩٩٣ [٢٥] الأسس الفنية لبناء مكنز عربى فى مجال تكنولوجيا التعليم. وهو يقترح أن يكون المكنز ثلاثى اللغة: عربى، إنجليزى، فرنسى. وهو لم يقدم مكنزاً وإنما قدم الأسس اللازمة لبناء المكنز والجديد هنا أن الباحث أعد إستبياناً للإفادة برأى الخبراء فى إجازة الأسس التى توصل إليها من دراسته للكتابات المرتبطة بإعداد المكانز بالإضافة إلى دراسة المكانز المنشورة.

وكان هؤلاء الخبراء فى مجالى المعلومات وتكنولوجيا التعليم.

على أن أهم عمل عربى فى هذا المجال هو مكنز الفيصل[٢٦] فقد اضطلع مركز الملك فيصل للبحوث والدراسات الإسلامية بمهمة وضع مكنز الفيصل للعلوم

(٢٥) على عبد الرحمن خليفة. أسس بناء مكنز عربى فى تكنولوجيا التعليم. ـ القاهرة، ١٩٩٣ (أطروحة ماجستير ـ جامعة حلوان. كلية التربية. قسم تكنولوجيا التعليم).

(٢٦) مركز الملك فيصل للبحوث والدراسات الإسلامية. الفيصل: مكنز عربى شامل فى علوم الحضارة: قسم علوم الدين الإسلامى. - الرياض: المركز، ١٩٩٤.- ٢ مج.

والمعارف الإسلامية نابع من الأسس التي تسير عليها تلك العلوم والمعارف لا على أساس الترجمة ومن هنا تميز هذا المكنز في أنه استند إلى المصطلحات العربية المنتزعة من واقع الإنتاج الفكري في العلوم الإسلامية فضلاً عن المصطلحات التي يستخدمها الباحثون في طلب موضوعاتهم.

وقد اشتمل المكنز على ٣١١١ مصطلحات كما احتوى على ٤٩٢٢ من اللاواصفات. ولا يقتصر مكنز الفيصل على العرض الهجائي المفصل للمصطلحات فحسب بل اشتمل على عرضين آخرين هما العرض المصنف والعرض التباديلي. وهكذا أتاح المكنز تعدد مداخل الوصول للمصطلح عن طريق عروض المصطلحات المختلفة.

ومن الأعمال الحديثة التي اتجهت نحو شمولية التغطية لموضوعات متنوعة نجد المكنز الموسع الذي صدر عام ١٩٩٦ [٢٧]. وهو ثلاثي اللغة: عربي ــ إنجليزي ــ فرنسي، ويشمل جميع ميادين المعرفة ويحتوي على ٢٤٣٨٢ واصفة و١٢٩٢٤١ لا واصفة باللغة العربية بالإضافة إلى ٨٥٧٠ لا واصفة بالإنجليزية و٢١٥٩ لا واصفة بالفرنسية. ويتألف المكنز من ثلاثة أقسام:

يتضمن القسم الأول جميع الواصفات بالعربية في تسلسل هجائي واحد ومع كل واصفة مقابلها باللغة الانجليزية والفرنسية، رمز الوجه، التبصرة، اللاواصفات، المصطلح الأعم، المصطلح الأضيق، المصطلح المترابط.

والقسم الثاني عبارة من قائمة الأوجه التي تشتمل على سبعة وعشرين وجهاً رئيسياً. وقد تم تفريع كل وجه رئيسي بدرجة من التفصيل اللازمة للموضوعات المعنية.

أما القسم الثالث فهو قسم الكشافات وتشتمل على كشافين هما الكشاف المصنف الذي رتب حسب رموز الأوجه وتحت كل رمز أدرجت جميع الواصفات في ترتيب هجائي، والكشاف الآخر هو كشاف الكلمات المفتاحية خارج السياق، وهو كشاف للكلمات المفتاحية الواردة في كل من الواصفات واللاواصفات مرتبة ترتيباً هجائياً

(٢٧) المكنز الموسع/ جمع وتحرير مُحمود أحمد اتيم. – عمان: مؤسسة عبد الحميد شومان؛ دبي: مركز جمعة الماجد للثقافة والتراث، ١٩٩٦. – ٣ مج (٢٧١٦ص)

حسب موقعها فى الألفبائية العربية وتحت كل كلمة يسرد الكشاف جميع الواصفات واللاواصفات التى وردت فيها تلك الكلمة المفتاحية ويكون مع كل واصفة رمز الوجه الخاص بها.

٧/٥ المكانز العربية تحت عين المجهر :

هناك العديد من الدراسات العربية التى تصف الجهود العربية التى تمت سواء فى إعداد المكانز أو فى تعريبها، ولكن توجد ندرة فى الدراسات التحليلية التقييمية للمكانز العربية سواء من وجهة نظر المعايير أو من وجهة نظر الاستخدام الفعلى لها.

دراسة واحدة تستحق الاعتبار هنا هى درسالة الماجستير لسوزان مصطفى فلمبان بعنوان: المكانز كأدوات للتحليل الموضوعى: دراسة تقويمية لاثنين من المكانز العربية [٢٨]. والمكنزان اللذان جرت دراستهما هما: مكنز مصطلحات علم المكتبات والمعلومات ومكنز الفيصل.

وقد جرى التقييم لهما من حيث نوعية الضبط المصطلحى والاستخدام الفعلى للمصطلحات من الواصفات واللاواصفات وحالة المصطلحات وحاجتها إلى تغييرات بالإضافة والتعديل عن طريق الكشف عن مدى التزام هذين المكنزين بالمواصفات التى وضعت لإعداد المكانز العربية واختبارها بشكل تطبيقى على عينة من الوثائق.

وهكذا فإن الدراسة تهدف إلى التأكد من مدى المطابقة للمواصفات العربية الخاصة ببناء المكانز وذلك من خلال وضع المكانز مجال الدراسة تحت ظروف الاستخدام الفعلى من جهة ومقارنتها بالمواصفة القياسية من جهة أخرى، وقد تم تناول المكنزين من حيث النقاط التالية:

- عرض وترتيب المكنزين
- المصطلحات
- علامات الترقيم وطرق تلافى الغموض

(٢٨) سوزان مصطفى فلمبان. المكانز كأدوات للتحليل الموضوعى: دراسة تقويمية لاثنين من المكانز العربية. - جدة ، ١٩٩٥. (أطروحة ماجستير ــ جامعة الملك عبد العزيز. كلية الآداب. قسم المكتبات والمعلومات).

ـــ العلاقات بين المصطلحات

ـــ طريقة البناء

ـــ الصيانة والتطوير

ثم قامت الباحثة باختيار عينة من الوثائق الصادرة بالعربية وفى المجالات الموضوعية التى يغطيها المكنزين لاختبار مدى كفائتهما عملياً فى ضوء التحليل الموضوعى للوثائق المختارة كعينة.

وقد توصلت الباحثة إلى بعض النتائج المهمة منها أن المكنزين لم يلتزما بجميع العناصر الواردة فى المواصفة القياسية العربية رقم (٥٧٨) لاعداد وتطوير المكانز رغم أن مكنز مصطلحات علم المكتبات والمعلومات قد نشر عام ١٩٨٠ أى قبل نشر المواصفة القياسية بأربع سنوات. وقدمت الباحثة فى نهاية دراستها بعض التوصيات، منها التخلى عن عملية إعداد المكانز العربية من واقع الأصول الأجنبية، وأن يتم بناء المكانز العربية كجهود إنشائية وفق الطريقة التحليلية للمصنفات واستفسارات الباحثين، وأنه من الضرورى الالتزام بالمنهجية المقررة فى المواصفات القياسية العربية وتطبيقها عند إعداد المكانز وتطويرها، وأنه من الواجب الاستفادة من البرامج الآلية الجاهزة عند بناء المكانز العربية.

۞۞۞

وهكذا فقد كانت المكانز العربية الأولى هى مجرد ترجمة أو تعريب لمكانز أجنبية، إلا أن الأعمال العربية الأخيرة هى أعمال إنشائية بالدرجة الأولى يجب تشجيعها والسير فى نفس الاتجاه بالنسبة للأعمال الجديدة. على أنه من الضرورى أن ننوه إلى ضرورة تبنى المفاهيم والأفكار الجديدة عند إنشاء مكانز جديدة أو تطوير مكانز موجودة، وضرورة التحديث المستمر للمكانز فليس الأمر هو إعداد مكنز بل أن يستمر هذا المكنز فى الاستخدام وفقاً لأحدث التطورات. ونشير أخيراً إلى ضرورة تطوير وتحديث المواصفات القياسية العربية الخاصة بإعداد المكانز وتطويرها.

الفصل الحادى العاشر

الاتجاه المكنزى فى قائمة رؤوس موضوعات مكتبة الكونجرس: ترجمة لمقدمة الطبعة الثالثة والعشرين لعام ٢٠٠٠م

مقدمة معد الترجمة:

يعتبر مُعدّ هذه الترجمة التى بين أيدينا هو أول من استخدم "مفهوم" المكنز Thesaurus فى اللغة العربية حين ترجم المصطلح إلى قاموس أو معجم ألفاظ أو مصطلحات وذكر كلمة Thesaurus الإنجليزية إلى جانب الترجمة العربية المذكورة، وكان ذلك عام ١٩٦٤[1] ولكن أول مقال باللغة العربية عن المكانز ويحتوى فى عنوانه على كلمة "مكانز" كان بقلم محمد فتحى عبد الهادى عام ١٩٧٨، وهو العالم العربى المصنف رقم (١) فى هذا الموضوع[2].

أما بالنسبة لترجمة الطبعة الثالثة والعشرين لقائمة رؤوس موضوعات مكتبة الكونجرس والتى صدرت عام ٢٠٠٠م، فتأتى كضرورة لتعريف الطلاب بآخر ما وصلت إليه الدراسات بالنسبة لأشهر قائمة رؤوس موضوعات فى العالم.. وهناك أيضا سبب راودنى حين بدأت الترجمة، وهو إحياء لذكرى رائد من الرواد العرب فى المجال وهو المرحوم/ إبراهيم الخازندار الذى تخرج فى قسم المكتبات والوثائق بجامعة القاهرة عام ١٩٥٧، وكانت له دراسة عن رؤوس الموضوعات العربية كمشروع بحث بالسنة الرابعة، وحازت هذه الدراسة إعجاب أساتذته وزملائه، واستمر اهتمامه بإعداد قائمة رؤوس الموضوعات العربية بعد ذلك، ولكنى أعتبر أن اختيارى له للعمل معى بجامعة الكويت، كرئيس لقسم الفهارس العربية منذ عام ١٩٦٨ حتى توفاه الله عام ١٩٩٣، كان مكسبا عربيا وليس كويتيا فحسب، ذلك لأننى قلت له: أنت رئيس لهذا القسم.. ولكن أرجو أن تركز جهودك فى إعداد قائمة

---

(١) أحمد بدر، التوثيق الآلى، ثورة فى عالم المكتبات.. مجلة المكتبة العربية القاهرة مج١، ع٤ (١٩٦٤) ص ١٨- ٢٦.

(٢) محمد فتحى عبد الهادى. المكانز كأدوات للتكشيف واسترجاع المعلومات، القاهرة: مكتبة غريب، ص ١٨

رؤوس موضوعات عربية يحتاجها العالم العربى، وستتولى السيدة/ سهير محفوظ(*) أعمال القسم وترجع إليك فقط عند وجود صعوبات ولاختبار القائمة بهذه الصعوبات من حين إلى آخر.. وصدرت قائمة الخازندار فى طبعة مبدئية قدمت للمؤتمر البيليوجرافى الأولى الذى عقد بالرياض عام ١٩٧٢، ثم توالت الطبعات والإصدارات، وكانت قائمة الخازندار منطلقا لدراسات وقوائم أخرى لرؤوس الموضوعات العربية.. تحية للرائد العربى. ودعوة من المكتبيين فى مختلف أنحاء الوطن العربى أن يتغمده الله برحمته.. وإذا كان الاتجاه المكنزى هو التطور الواضح فى قوائم رؤوس الموضوعات وعلى رأسها قائمة مكتبة الكونجرس، فلعل المنظرين والممارسين العرب يرون فى هذه الترجمة ما يحفز هممهم لبذل جهود عربية مخلصة لمسايرة هذا التطور.

وعلى الرغم من أن تنظيم المعرفة أساسى لمهنة المكتبات، إلا أن البعض قد يشعرون بأن التطور فى تكنولوجيا المعلومات سيقلل من أهمية المقررات المحورية فى التحليل الموضوعى، إلا أن الحقيقة المعاصرة، تشير إلى استمرار أهمية التحليل الموضوعى فى أدوات الاسترجاع، سواء بالنسبة للأوباك OPAC أو قواعد البيانات البيليوجرافية أو نظم النص الكامل أو الإنترنت (٤)

مقدمة الطبعة :

تحتوى الطبعة الثالثة والعشرون من قائمة رؤوس موضوعات مكتبة الكونحرس (LCSH 23) على الرؤوس التى وضعتها المكتبة حتى ديسمبر ١٩٩٩، وقد تم إعداد هذه القائمة عن طريق إنشاء شريط إليكترونى Tape يحتوى على جميع الرؤوس التى تم التحقق منها فى الملف المرجعى للموضوعات بمكتبة الكونجرس حتى يناير ١٩٩٩، ويلاحظ أن هذا الملف يحتوى على حوالى (٢٥١,٣٠٠) ألف تسجيلة مرجعية بالمقارنة بحوالى (١٩٩) تسجيلة فى الطبعة السادسة عشرة لقائمة مكتبة الكونجرس، التى سبق للكاتب ترجمتها.

(*) دكتورة سهير محفوظ تعمل حاليا رئيس لقسم المكتبات والمعلومات بجامعة حلوان.

(4) Williamson, N. J. (1997) The Importance of subject Analysis in library and Information Science Education. **Technical Services Quarterly**, V. 15 (1/2), p. 67-86

وتحليل قاعدة البيانات الخاصة بالمرجعية الموضوعية للرؤوس فى الطبعة الثالثة والعشرين، يدلنا على أن الملف يحتوى على ١٨,٨٠٠ اسم شخص (كرؤوس موضوعات) من بينها ١١٣٠٠ مدخل تمثل أسماء العائلة، بالإضافة إلى (٤٣٠٠) رأس مشارك وأقل من ١٢ رأس موضوع للاجتماعات والمؤتمرات و٤٨٠ عنوان موحد Uniform titles بالإضافة إلى ١٨٩,٠٠٠ رأس موضوع عام Topical بالإضافة إلى ٣٩,٠٠٠ رأس جغرافى، كما يقدر احتواء هذه القائمة على ٦٠٠ إحالة استخدم use العامة، وكذلك ٣٦٠٠ إحالة انظر أيضا العامة، وهناك ٢٢١,٠٠٠ إحالة من رأس موضوع غير مستخدم إلى رأس موضوع مستخدم.

إن خلق أو إنشاء وتغيير رؤوس الموضوعات هى عملية مستمره حيث يضاف من ٦٠٠٠- ٨٠٠٠ رأس موضوع إلى القائمة كل عام، بما فى ذلك رؤوس الموضوعات ذات التقسيمات الفرعية Sub divisions، ويلاحظ أن الاقتراحات بإضافة أو تغيير رؤوس موضوعات تقدم إلى المفهرسين بمكتبة الكونجرس أو إلى المفهرسين فى المكتبات التى تشترك مع مكتبة الكونجرس فى برنامج تعاونى رسمى، والاقتراحات التى تتم الموافقة عليها تصبح جزءا من الملف المرجعى على الخط المباشر On- Line authority file لقائمة رؤوس موضوعات مكتبة الكونجرس، والتى يمكن عن طريق الملف إصدار أو إعداد مطبوعات مختلفة.

وهناك خمس خدمات تقدم المعلومات الخاصة برؤوس الموضوعات الجديدة أو التى تغيرت وهى كما يلى:

١- خدمة توزيع أشرطة الحاسبات: حيث توفر رؤوس الموضوعات فى شكل مارك الأمريكى MARC 21 والذى يصدر أسبوعيا، وذلك لاستكمال ملف قاعدة البيانات الأساسى الخاص بالتسجيلات المرجعية للموضوعات.

٢- القائمة الأسبوعية لقوائم رؤوس الموضوعات الدالة على رؤوس الموضوعات الجديدة أو التى تغيرت بالإضافة إلى أرقام التصنيف والإحالات والملاحظات الخاصة بالحواشى Scope Note وهذه القوائم تحمل اسبوعياً على الوب (WWW) بالعنوان التالى:

< URL. http: // IC web: Loc. gov/ catdir/ epso/ wls. html>

٣- قائمة رؤوس الموضوعات على هيئة ميكروفيش وتصدر أربع مرات فى العام حيث تعتبر ممثلة لطبعة جديدة من قائمة رؤوس الموضوعات كل ثلاثة شهور.

٤- يوزع الملف المرجعى الكامل أربع مرات سنوياً على هيئة أقراص مكتنزة CD-ROM وهذه توزع تحت عنوان: Classification PLus

٥- نشرة خدمة الفهرسة: حيث تقدم لنا المعلومات الخاصة برؤوس الموضوعات الجديدة والتى تغيرت أو ألغيت.

أولاً: استخدام القائمة مع الأدوات المساعدة:

يجب استخدام الطبعة الثالثة والعشرين مع أدوات مساعدة عديدة وأهمها دليل الفهرسة الموضوعية: رؤوس الموضوعات (الطبعة الخامسة ١٩٩٦).

Subject Cataloging Manual: Subject Headings (5th edition, 1996).

ويتم الحصول على هذا الدليل عن طريق الاشتراك لتحديث الخدمات، ويحتوى الدليل على نفس التعليمات المستخدمة بواسطة المفهرسين الموضوعيين بمكتبة الكونجرس فى عملهم اليومى، وعلى الرغم من أن بعض هذه التعليمات تشرح الإجراءات الداخلية لمكتبة الكونجرس، إلا أن معظم هذه التعليمات أساسية لأولئك الذين يرغبون فى فهم وتطبيق قائمة رؤوس موضوعات مكتبة الكونجرس بطريقة صحيحة ويتم الإحالة للدليل إذا كانت هناك معلومات إضافية عن موضوع معين مشروحة فى هذا الدليل، ويلاحظ أن نشرة خدمة الفهرسة تحتوى على المعلومات الخاصة بأهم الموضوعات التى تغيرت منذ الطبعة السابقة للدليل، كما تحتوى النشرة على قوائم رؤوس الموضوعات ذات الاهتمام الجارى، وكذلك قوائم الرؤوس التى تغيرت حديثاً فضلا عن المعلومات الخاصة بنشرات رؤوس الموضوعات، هذا والرؤوس الخاصة بأسماء الأشخاص أو الهيئات أو التشريعات أو غيرها من الرؤوس التى يمكن تقليديا أن تكون مكان المؤلف، هذه الرؤوس يتم وضعها كرؤوس موضوعات بواسطة المفهرسين الموضوعيين والتسجيلات المرجعية التى تضم هذه الرؤوس موجودة فى المطبوعات التالية:

(أ) طبعة الميكروفيش التركيمية المرجعية للأسماء.

Name Authorities Cumulative Microfiche Edition.

(ب) أسماء مارك على الأقراص المكتنزة CD MARC Namesوهذه يجب استشارتها للتعرف على الشكل الموافق عليه.

وعندما تكون رؤوس الأسماء مطبوعة فى قائمة رؤوس موضوعات مكتبة الكونجرس "Library of Congress Subject Headings" "LCSH"

فمعنى ذلك أنه قد تم استعارتها من الملف المرجعى للأسماء، أما البناء المرجعى الكامل والمعلومات المرجعية الإضافية فستظهر فقط فى الملف المرجعى للأسماء.

التعريفات الحرة العائمة: Free- Floating

ينشر سنوياً كشاف هجائى يحتوى على جميع التعريفات الحرة العائمة والتى تظهر تحت الكود H 1095 -H 1200 فى دليل الفهرسة الموضوعية: رؤوس الموضوعات، هذا يوجد تحت كل تفريع قائمة بالرؤوس الحرة العائمة حسب فئاتها Categories كما يوجد إيضاحات للتفريعات التى لها أوراق تعلميات محددة فى الدليل حيث يتم وصف استخداماتها، بالإضافة على الحقل الفرعى كود مارك أى أن القائمة يجب أن تستخدم مع الدليل. ويمكن الحصول على مطبوعات رؤوس الموضوعات من:

Cataloging Distribution Service

Library of Congress

Washington, D. C. 20541- 5210.

Internet: eds info @ Loc. gov.

أما الأسئلة التى قد تثار حول محتوى هذه المطبوعات فتوجه إلى:

Cataloging Policy and Support Office.

Library of Congress

Washington, D. C. 20540- 4305.

Internet : cpso @ Loc. gov.

ثانياً : البدايات التاريخية:

تعتبر الطبعة ٢٣ مثل سابقاتها تركيما لرؤوس الموضوعات التى بدأتها مكتبة الكونجرس منذ عام ١٨٩٨، وفى تلك السنة قدرت المكتبة ضرورة توفير فهرس قاموسى بدلاً من "الفهرس المصنف الهجائى" أو "الفهرس المصنف"، وذلك للتكامل مع نظام التصنيف الجديد الذى حل محل نظام Jefferson، وقد أعدت قائمة لرؤوس الموضوعات المستخدمة فى الفهرس القاموسى بواسطة لجنة من جمعية المكتبات الأمريكية ونشرت عام ١٨٩٥ وتعتبر هذه أساس رؤوس الموضوعات الجديدة بالمكتبة. وكانت أول طبعة من قائمة مكتبة الكونجرس تسمى:

"رؤوس الموضوعات المستخدمة فى الفهارس القاموسية لمكتبة الكونجرس" "وقد طبعت هذه القائمة فى أجزاء بين عامى ١٩٠٩- ١٩١٤، ثم صدرت قوائم، تكميلية، ولكن الطبعة الثانية صدرت لتجميع هذه القوائم فى عام ١٩١٩، أما الطبعات التالية فقد نشرت فى فترات غير منتظمة، كما تغير عنوان القائمة منذ الطبعة الثامنة المنشورة عام ١٩٧٥ إلى "رؤوس موضوعات مكتبة الكونجرس" "LCSH".

وقد تطورت واتسعت هذه القائمة عبر الزمن لتعكس الفلسفات المتباينة للمكتبة، ولمئات المفهرسين الذى أسهموا فى وضع رؤوس الموضوعات، ولعل عدم الإنتظامات فى صياغة بعض رؤوس الموضوعات يمكن أن يرد إلى هذه السياسات والافراد المختلفين.

مكونات مداخل رؤوس الموضوعات:

تكتب الرؤوس المستخدمة بحروف بارزة سوداء مثل Nuclear Physics ويمكن أن يكون هذا الرأس متبوعاً بالجملة التالية بين قوسين (يمكن التقسيم الفرعى جغرافيا) وهذا يدل على أن رأس الموضوع يمكن تفريعه بواسطة المكان طبقاً للقواعد الموجودة فى الدليل "Manual"وقواعد أرقام التصنيف. أما ملاحظات الحاشية "Scope Note"والتى ترشد إلى المعنى أو تطبيقات رأس الموضوع، فيمكن أن تأتى فى فقرات متفرقة، والإحالات المرتبطة برؤوس الموضوعات ترتب فى مجموعات متبوعة بالتفريعات الخاصة برؤوس الموضوعات وهذه المكونات يمكن التعرف عليها فيما يلى:

رؤوس الموضوعات:

يمكن أن تتكون هذه الرؤوس من كلمة واحدة أو عدة كلمات، ورأس الموضوع المكون "من كلمة واحدة" يعتبر عادة اسما (مثل.. "Schools, Vis- " "cosity") وعادة تكون المفاهيم فى الصيغة المفردة أما الأشياء فتكون بصيغة الجمع، وهناك استثناءات قد توجد بالنسبة لذلك.

وعندما يكون رأس الموضوع "من كلمتين" فهذه تحتوى عادة على صفة واسم، وهذه قد تظهر فى ترتيب الكلمات العادى مثل "Nuclear Physics" أو قد تكون فى شكل مقلوب، ولكن الشكل المقلوب يكون مع الصفات التى تصف اللغة أو الجنسية مثل"Art French"أو "Songs, American" وهناك أشكال أخرى من الرؤوس التى يمكن قلبها حتى يكون الاسم فى البداية مثل "Bridges, Concrete"والسياسة الحالية هى استخدام نظام الكلمات العادية لرؤوس الموضوعات باستثناء الرؤوس المصحوبة باللغة أو الجنسية أو الصفات العرقية، وكذلك الرؤوس الموصوفة بالفترة الزمنية"Art, Medieval" أو الرؤوس الموصوفة موسيقياً أو فنياً.

وعلى الرغم من أن الهدف الأساسى هو اتباع رؤوس موضوعات فى الترتيب القاموسى بدلاً من الترتيب الهجائى المصنف. إلا أن القائمة تعكس مداخل منتشرة متعلقة بعضها ببعض، وهناك العديد من رؤوس الموضوعات التى وضعت فى الأساس حسب اسم الفئة أولاً "Class First" عن طريق استخدام التفريعات أو القلب أو وضعها بين قوسين وهذه الرؤوس استمرت فى القائمة حتى اليوم ومن أمثلتها:

"Railroads- Tickets" أى استخدام الشرطة.

"Cookery (Fish)" أى استخدام بالقوسين.

"Insurance, Fire" أى استخدام الفاصلة.

هذا، والأسماء ذات الشكل الجغرافى تقلب عادة لوضع الكلمة الدالة فى البداية وذلك مثل: Lake Erie حيث توضع تحت Eire, Lake حتى يظهر الجزء المميز وهو Erie فى البداية، وعند استخدام اكثر من كلمتين فى رأس الموضوع

فإن الرأس قد يشمل جملا أو كلمات توضيحية، أما الرؤوس التى تشمل كلمة " and"فيمكن أن يتم التعبير عنها بعلاقة تبادلية مثل:

"Technology and Civilization" كما أن الرؤوس التى تحتوى على جمل شارحة يمكن أن تقلب كما يلى:

"Plants, Effect of the moon on" أو أن تكون فى الترتيب العادى مثل: "Photography of birds" وإن كانت مكتبة الكونجرس قد قامت بتغيير بعض الرؤوس المقلوبة إلى رؤوس ذات تفريعات وذلك طبقاً لكل حالة على حدة.

أرقام التصنيف :

تتوافر أرقام التصنيف الخاصة بمكتبة الكونجرس فى حوالى ٣٦% من رؤوس الموضوعات، وأرقام التصنيف هذه تمثل اكثر الجوانب عمومية فى الموضوع، وإذا كانت هناك عدة جوانب للموضوع فيتم تغطيتها بعدة أرقام تصنيف مختلفة وذلك كما يلى:

Diesel MOTOR (TJ 759)
NORWEGIANLANGUAGE (PD 2571- PD 2699)
SHELLFISH [QL 401- QL 445 (ZOOLOGY)]
[RA 602, S6 (PUBLIC HEALTH)]
[TX 387 (SHELLFISH AS FOOD)]

ونظراً لان هذه الأرقام وكذلك رؤوس موضوعات مكتبة الكونجرس معرضة للتغيير والمراجعة بصفة مستمرة، فلا ينبغى استخدام أرقام التصنيف الموجودة فى قائمة الكونجرس إلا بعد التحقق منها فى الطبعات الأخيرة أو فى الملاحق.

ملاحظات الحواشى"Scope notes"

توجد ملاحظات الحواشى هذه عندما تكون ضرورية لضمان الانتظام فى استخدام رؤوس الموضوعات، وذلك عن طريق تحديد نطاق المادة الموضوعية التى يتناولها رأس الموضوع فى فهارس المكتبة حيث تدلنا هذه الملاحظات على التمييز الضرورى بين رؤوس الموضوعات ذات العلاقة ببعضها أو لبيان المعنى المقصود من رأس الموضوع والمستخدم فى فهارس المكتبة، وتظهر هذه

الملاحظات فى القائمة تابعة مباشرة لرؤوس الموضوعات، ويوجد فى قائمة مكتبة الكونجرس حوالى ٥٠٠٠ ملاحظة حواشى.

ثالثاً: الإحالات أو بيان العلاقات بين رؤوس الموضوعات:

تحتوى قائمة مكتبة الكونجرس على إحالات عديدة وضعت فى أوقات مختلفة، طبقاً لفلسفات مختلفة اتبعت فى هذه الأوقات، وقد ظلت بعض الإحالات من الموضوع المحدد للموضوع العام، كما هى منذ بداية هذا القرن، وقد وضعت قواعد جديدة منذ وقت قريب حيث تركز هذه الإحالات على العلاقات الهرمية، كما تم حذف العديد من الإحالات غير الدقيقة من القائمة، وقد تم شرح هذه القواعد بالتفصيل فى الدليل Manna.

علاقة التكافؤ: إحالات استخدام USE Reference

تستخدم هذه الإحالة "use" من مصطلح غير مستخدم أو غير مفضل إلى مصطلح مستخدم ومفضل، وتستخدم حروف (UF) (USED For) تحت رأس الموضوع الذى يتم الإحالة إليه وذلك كما يلى:

Cars (Automobiles)
USE Automobiles
Automobiles
UF Cars (Automobiles)

أى أن كلمة "USE" الحالية هى البديل لإحالة انظر التى سبق استخدامها فى الطبعات السابقة. أى أن الاختصار UF معناها استخدم بدلا منه. ويلاحظ أن كلمة USE أو الكود UF تظهر فقط أمام الإحاطة الأولى إذا كانت هناك عدة إحالات مثل:

Raw Foods
UF Food, Raw
Uncooked Food
Unfired Food.

أما إذا كان هناك شكل قديم لرأس موضوع فهناك إحالة بعد هذا الرأس يذكر فيها رأس موضوع سابق [Former heading].

مثال **On Line Catalogs**

UF Catalogs, on Line

Former heading

أى أن إحالة استخدم "USE" يتم عملها من المترادفات أو من الهجاء المختلف للكلمات، أو لأشكال مختلفة من التعبير أو للرؤوس المقلوبة، وكذلك بدلاً من رؤوس الموضوعات القديمة، كما تستخدم إحالة "USE" أيضا عندما يتقرر عدم استخدام كلمات معينة كرؤوس موضوعات، حتى إذا كانت هذه الرؤوس والكلمات غير المستخدمة ليست مترادفات، كما تستخدم رؤوس الموضوعات التى تحتوى على أكثر من كلمة واحدة إحالة"USE" من الكلمات غير المختارة كمدخل، كما أن إحالات USE لا تتم عادة فى القائمة من المختصرات، كما لا تتم من ترجمات اللغة الأجنبية، هذا وإحالات"USE" تلغى عادة إذا كانت تبدأ بنفس الكلمة كمصطلح عريض مطلوب فى التسلسل الهرمى مثال:

Exterior Lighting

BT = Broadterm

هذه الإحالة السابقة Exterior Lighting يتم عملها بدلا من الإحالة التالية

UF Lighting, Exterior.

وتظهر العلاقات الهرمية فى BT مصطلح اعرض وNT مصطلح أضيق، أى أن رؤوس الموضوعات يتم ربطها برؤوس موضوعات أخرى، من خلال كل من إحالة BTوNT ويسبق الكود "BT" رأس الموضوع المعبر عن رقم التصنيف. أما الكود "NT"فيسبق رأس الموضوع الذى يمثل فى معظم الحالات جزءا من رقم التصنيف الذى يمثله الرأس الذى يظهر تحته الكود NT.

وتعتبر الأكواد BT و NT كأكواد تبادلية، ذلك لأن رأس الموضوع الذى

يظهر به الرمز BT يجب أن يقابله العلاقة العكسية للكود NT كما هو المثال التالى:

Exterior Lighting
BT Lighting
Lighting
NT Exterior lighting

ويلاحظ أن رأس الموضوع يكون موصولا عادة برأس موضوع آخر يأتى بعده مباشرة فى التسلسل الهرمى لرؤوس الموضوعات، وبالتالى فالإحداث التى تتم معبرة عن العلاقات البعيدة لم تعد مستخدمة، أى أن الإحالات التى تؤدى إلى اثنين أو اكثر من المستويات فى التسلسل الهرمى تعكس ممارسة قديمة منتهية، والنتيجة التى نراها بعد عمل العلاقات الهرمية الواضحة تخلق نظاما واضحا من الأبوة والبنوة، ورؤوس الموضوعات التى تم وضعها بعد عام ١٩٨٤ يجب أن تتبع هذه المبادئ، أما رؤوس الموضوعات التى وضعت قبل عام ١٩٨٥ فتتم مراجعتها على أساس تدريجى حيث تتم التغييرات فى الإحالات لتناسب القواعد الجديدة، وحتى يتم الانتهاء من هذه المراجعة فستحتوى القائمة وللمستقبل القريب إحالات لا تعكس العلاقة الهرمية. إن وضع وإنشاء الإحالات الهرمية يخلق القدرة على العثور منهجياً على قوائم رؤوس الموضوعات الأكثر عمومية أو الأكثر خصوصية من رأس الموضوع الذى تتم استشارته. وعلى أى الحالات فمهما كان المستوى الذى يدخل به الباحث فى العلاقة الهرمية فالباحث يمكن أن يتبع إما المصطلحات الأعرض "BT" أو المصطلحات الأضيق أو الأكثر تحديدا "NT" للعثور على رؤوس الموضوعات العريضة، أو الأكثر تحديداً المتوافرة وذلك كما يظهر فى المثال التالى:

Vehicles
BT Transportation
NT Motor- Vehicles
Motor- Vehicles
BT Vehicles
NT Trucks

Trucks

BT Motor- Vehicles

NT Dump Trucks

Dump Trucks

BT Trucks.

وبالتالى فيوضح لنا الكود "NT" أن أكثر رؤوس الموضوعات تحديداً هى Dump Trucks أما باتباع الكود "BT" فمن الواضح أن أكثر رؤوس الموضوعات تحديداً فى الاتساع هو "Transportation" ويلاحظ أن هناك العديد من الإحالات الهرمية المستخدمة فيما سبق قد تم إلغاؤها عندما يكون رأس الموضوع بادئا بنفس الكلمة التى يبدأ بها رأس الموضوع الأعرض ولعى سبيل المثال: فإن مصطلح Bridges, Iron and steel لا يحتوى الكود BT Bridges: وذلك لأن الترتيب الهجائى لرأس الموضوع الذى يترتب بعده مباشرة، كما يلاحظ أن رؤوس الموضوعات الأعرض تتم إضافتها تدريجياً كلما دعت الحاجة إلى ذلك بواسطة اتباع قواعد التسلسل الهرمى دون اعتبار للترتيب الهجائى.

العلاقات الترابطية: **The Associative Relationship**

يعبر عن هذه العلاقة الترابطية أو علاقات الاقتران بالكود.

Related Topics ◄——— RT"

Or أو Related Term

حيث يربط هذا الكود بين رأسين من رؤوس الموضوعات لهما اقتران بطريقة ما غير الطريقة الهرمية، وذلك مثل:

أى له علاقة ترابطية Ornithology والعكس RT Birds

ولكن السياسة العامة الحالية فى مكتبة الكونجرس هى ربط رؤوس موضوعات قليلة كإحالات ترابطية، وذلك حتى تتم مراجعة القوائم بالنسبة للعلاقات الهرمية بدقة.

الإحالات العامة:

الإحالة العامة هى الإحالة التى تتم ليس إلى رؤوس موضوعات محددة، بل إلى مجموعة كاملة من الرؤوس، وقد كانت الممارسة السابقة تعتبر قوائم الإحالات المخصصة لجميع رؤوس الموضوعات التى يحتويها رأس موضوع عريض، هذه الممارسة غير عملية، وإن كانت الإحالات المخصصة هذه تعتبر من الناحية النظرية صحيحة ومنطقية، وخصوصاً أن هذه الرؤوس المنفردة كانت مطبوعة فى قائمة الكونجرس، فبدلا من ذلك فإن إحالة انظر أيضاً العامة قد تم استخدامها مع الاحتفاظ بالكود SA (See also).

مثال :

Gums and resins ◄——— من الإحالة العامة إلى

SA Specific gums and Resins

e.g. Copal, Kaurigum, Kino ———► الإحالة المخصصة.

ومن المتوقع أن تقوم كل مكتبة بعمل إحالات مخصصة لكل واحد من هذه المصطلحات المخصصة مثال: (e. g. Copal) والتى ستوجد بالمكتبة. ومعنى ذلك أن العديد من الإحالات العامة سيتم استبدالها تدريجياً بالإحالات المخصصة لرؤوس الموضوعات الأضيق.

كما يلاحظ أيضاً أن الإحالات العامة قد تم عملها من رؤوس الموضوعات العامة إلى مجموعة رؤوس الموضوعات التى تبدأ بنفس الكلمة.

Chemistry: **مثال**

SA headings with the
Word Chemistry

وهناك أيضاً بعض الإحالات العامة التى تؤدى إلى تقسيمات فرعية.

Economic history :مثال

SA Subdivision Economic
Conditions under names of countries, cities.. etc.

أى انظر أيضا التقسيمات الفرعية، الأحوال الاقتصادية تحت أسماء البلاد أو الأقطار .. الخ.

أى أن إحالة انظر أيضاً تتم إلى التقسيم الفرعى: الأحوال الاقتصادية تحت أسماء البلاد، المدن... الخ.

وإحالة استخدم العامة "use" يتم عملها كما يلى:

Access Control
USE Subdivision
Access Control Under Subjects:
e. g. Computers- Access Control
Psychiatric records- Access Control.

أى أنه تم استخدام التقسيم الفرعى "التحكم فى الوصول" تحت الموضوعات كما يلى:

الحاسبات الآلية- التحكم فى الوصول.
التسجيلات النفسية- التحكم فى الوصول.

ومثل هذه الإحالات العامة إلى التقسيمات الفرعية لا تدل بالضرورة على أن هذه التقسيمات الفرعية مصطلحات حرة عائمة Free- Floating.

التقسيمات الفرعية

يتطلب تطبيق رؤوس موضوعات مكتبة الكونجرس الاستخدام الواسع للتقسيمات الفرعية الموضوعية، كوسيلة لضم عدد من المفاهيم المختلفة داخل رأس موضوع واحد، والموضوعات المعقدة يمكن تمثيلها بواسطة رؤوس موضوعات متبوعة برؤوس فرعية، وبعض هذه التفريعات مطبوعة فى الطبعة ٢٣ من القائمة، ولكن عدداً أكبر من هذه التفريعات يمكن أن توضع طبقاً لقواعد محددة فى دليل العمل، أى أن جزاءاً فقط من التوليفات أو التعليقات الخاصة بالرؤوس والتفريعات هى التى توجد فى قائمة رؤوس موضوعات الكونجرس، وحتى تسهل قراءة القائمة فالأجزاء الأولى من رأس الموضوع مع التفريعات لا تظهر فى

الطباعة، وبدلاً من ذلك تظهر التفريعات فى القائمة تالية لشرطة طويلة دون تكرار لرأس الموضوع مثال:

Massachusetts
Antiquities

وهذا المثال السابق تم إنتاجه من تسجيلة مقروءة آلياً حسب الوضع التالى:

Mass a chusetts .... Antiquities

وفى حالة استخدام اثنين من التفريعات فإن رأس الموضوع الأساسى والتفريع الأولى يستبدلان بشرطتين طويلتين مثل:

Mass a chusetts
Description and Travel
...1951- 1980
...1981-

وهذه البيانات محمولة على التسجيلات المرجعية الموضوعية كما يلى:

Massachusetts Description and Travel- 1951- 1980 and Massachusetts Description and Travel- 1981-

أى أنه عند وجود رأس موضوع له عدة تفريعات يتم تفريعها أيضا.

مثل: Untied States - History

كما أن الشرطة (—) تساعد فى تحديد التقسيمات الفرعية بطريقة صحيحة.

فئات التفريعات:

هناك أربع فئات من التفريعات يمكن التعرف عليها بصفة عامة كما يلى:

تفريع موضوعى: Topical
تفريع بالشكل: Form
تفريع زمنى: Chronological
تفريع جغرافى: Geographic

وكل واحدة من هذه الفئات يمكن شرحها فى السطور التالية، وهناك أمثلة عديدة لها فى قائمة رؤوس موضوعات الكونجرس فضلا عن وجود التعليمات الخاصة بكيفية تحديد ووضع هذه الفئات التى تظهر فى أقسام مختلفة من دليل العمل Manual.

التفريعات الموضوعية:

التفريعات الموضوعية تستخدم تحت رؤوس الموضوعات الرئيسية أو تحت التفريعات لتحديد المفهوم الذى يتم التعبير عنه من رأس الموضوع الفرعى المتخصص، على سبيل المثال:

Corn- Harvesting Automobiles
Motors- Carburetors and
Women- Employment

ويلاحظ أن العديد من التفريعات الموضوعية تحذف من القائمة المطبوعة، ولكن قواعد استخدامها توجد فى دليل العمل، وكذلك توجد فى الإحالات العامة المطبوعة تحت رؤوس الموضوعات الرئيسية فى قائمة رؤوس الموضوعات الخاصة بالكونجرس (LCSH).

تفريعات الشكل:

تفريعات الشكل تستخدم لبيان الشكل الذى تنظم وتقدم به المادة عن موضوع معين، والأشكال تشمل: مؤتمرات Congresses أو قواميس Dictionaries أو دوريات Periodicals .. إلخ، والشكل فى هذه الحالة يضاف للعنصر الأخير من أى رأس موضوع، وتفريعات الشكل تمثل شكل العمل وليس موضوعه. وبالتالى فهو يستخدم عادة تحت أى موضوع، من أجل ذلك فمن النادر طبع هذا التفريع فى قوائم رؤوس الموضوعات، مع ذلك فهناك بعض الحالات فى القائمة وخصوصاً تلك الرؤوس التى وضعت قبل عام ١٩٧٤. هذه الرؤوس أصبحت بعد ذلك حرة وعائمة على سبيل المثال:

Massachusetts - History - Colonial
Period 1600 - 1776 - Juvenile Literature

Unitd States - History - Periodicals

هذا ومعظم تفريعات الشكل تظهر فى القائمة بعد الإحالة العامة "انظر أيضاً" See also تحت رأس الموضوع الذى يمثل الشكل ككل. مثال: الدوريات

Periodicals

SA Subdivition Periodicals

Under Specific Subjects e.g.

Engineering - Periodicals

United States - History - Periodicals.

وباللغة العربية يمكن أن تكون كما يلى:

الدوريات

انظر أيضاً تفريع دوريات تحت رؤوس الموضوعات المخصصة على سبيل المثال:

الهندسة – الدوريات.

الولايات المتحدة – تاريخ – دوريات.

هذا والتعليمات والتوجيهات الخاصة بالاستخدامات المتصلة بتفريعات الأشكال العديدة مثل المستخلصات Abstracts والفهارس Catalogs والقواميس Dictionaries وكتب الحقائق Handbooks والجداول Tables وغيرها، هذه موجودة فى دليل العمل.

التفريعات الزمنية: Chronological Subdivisions

تستخدم التفريعات الزمنية لتحديد رؤوس الموضوعات أو تحديد الرؤوس وتفريعاتها بالنسبة لفترة زمنية معينة وتحت أسماء البلاد أو الأقاليم، حيث تطبع التفريعات الموضوعية أو الزمنية المحددة التى يمكن استخدامها مع أسماء هذه البلاد أو المناطق، ويوضع فى قائمة رؤوس الموضوعات تفريعات زمنية كنماذج وأمثلة فقط. وذلك كما يلى:

الولايات المتحدة – الأحوال الاقتصادية

United States - Economic Conditions

United States - History

United States - Politics and Government

الولايات المتحدة – السياسة والحكومة.

وعندما تحتوى الرؤوس الموضوعية على تفريعات زمنية غير مسبوقة بالتفريع – History تاريخ، فإن هذه التفريعات تنشأ وتطبع فى قائمة رؤوس الموضوعات وذلك كما يلى:

Philosophy, French - 18th Century

الفلسفة الفرنسية – القرن ١٨.

Art, Chinese - To 221 B.C

الفن الصينى – إلى ٢٢١ ق.م.

التفريعات الجغرافية:

هناك نوعان من الحاشية التى يمكن تفريعها جغرافيا بعد رأس الموضوع أو بعد التفريعات حيث تبين لنا أن المنطقة الجغرافية يمكن أن تتبع رأس الموضوع أو التفريعات أما الحاشية (التى لا يتم تفريعها جغرافيا) بعد رأس الموضوع أو التفريع تدلنا على أن هناك قراراً قد اتخذ لعدم تقسيم رأس معين حسب المنطقة الجغرافية، أما حذف الحاشيتين أعلاه فيعنى عادة أن هذا الرأس لم يتم اتخاذ قرار بشأنه من ناحية إمكانية التفريع الجغرافى وبالتالى فالمنطقة الجغرافية لا يجب استخدامها.

والتعليمات المتعلقة بتفريع بالمكان يمكن أن تتم تحت رأس الموضوع، مع ملاحظة الحاشية (SCOPE NOTE)، ولكن الوصف الكامل للقواعد المستخدمة يوجد فى الدليل، وبصفة عامة إذا كان الكيان الجغرافى هو اسم للدولة أو اسم أكبر من دولة واحدة، فإن الاسم يوضع مباشرة بعد الرأس أو التفريع الذى يحتوى على ملاحظة الحاشية (يمكن تفريعها جغرافياً) فمثلاً:

LABOR SUPPLY (MAY SUBD GEOG)

أى تزويد العمالة (يمكن تفريعها جغرافيا)، وهذا يعنى أن المكان يتبع الموضوع مثل: LABOR SUPPLY - France. وإذا كان الكيان الجغرافى هى

اسم ولاية أو مدينة، فإن البلد تسبق اسم المنطقة الجغرافية الأصغر، أى أننا بهذه الممارسة نجمع المناطق المحلية أو المدن كتفريعات تحت اسم الدولة.

مثل: Labor Supply - France-Paris

العمالة ـــ فرنسا ـــ باريس.

والاستثناء الأساسى من هذه القاعدة الخاصة بوضع الأقطار بين الموضوع والبلد المحدد يتم بالنسبة لثلاثة أقطار هى "الولايات المتحدة - بريطانيا - كندا".

أى أننا فى هذه الحالة سوف نستخدم اسم الولاية أو الإقليم أو المحافظة أو المناطق باعتبارها المجمعة للكيانات الجغرافية الأصغر، وهناك استثناءات أخرى للقاعدة العامة الموضحة فيما سبق، وهذه مشروحة فى دليل العمل Manual .

وإذا كان الرأس يحتوى على كل من التفريع الجغرافى والموضوعى والشكلى، فإن التفريع الجغرافى يعتمد على أى العناصر يمكن تفريعها بواسطة المكان، ولكن كقاعدة عامة فالمكان يتبع العنصر الأخير الذى يمكن تقسيمه بالمكان وذلك مثل التفريعات التالية:

Construction industry (May subd Geog)

صناعة البناء (يمكن تفريعها جغرافيا).

–Finance التمويل

- Law and legislaion (May Subd. Geog)

– القانون والتشريعات (يمكن تفريعها جغرافيا).

- Government policy (May Subd Geog)

– السياسة الحكومية (يمكن تفريعها جغرافيا)

- Mathematical Models

– النماذج الرياضية وهذه الرؤوس ستؤدى إلى التوليفات التالية:

Construction - Italy

Construction Industry - Italy - Finance

Construction Industry - Finance. – Law and Legislation-Italy

Construction Industry - Government Policy - Italy.

Construction Industry - Italy – Mathematical Models.

وترجمتها بالعربية كما يلى:

البناء - إيطاليا

صناعة البناء - إيطاليا - التمويل

صناعة البناء - التمويل - قانون وتشريع - إيطاليا

صناعة البناء - سياسة حكومية - إيطاليا

صناعة البناء - إيطاليا - النماذج الرياضية

ويلاحظ أن هناك نماذج قليلة من التفريعات الجغرافية الموجودة فى قائمة رؤوس الموضوعات على سبيل المثال رأس موضوع.

Petroleum wastes (May subd geog)

أى يكون متبوعاً بالحاشية (يمكن تفريعها جغرافيا) ولكن لا يتم وضع أو ذكر بلاد معينة، وفى هذه الحالة فالقواعد المحددة للتفريع الجغرافى فى الدليل يجب استشارتها لتركيب رأس الموضوع الصحيح بالتفريعات الجغرافية السليمة، وهذه التفريعات الجغرافية المطبوعة فعلا فى القائمة، تكون تفريعات مطلوبة، حتى تتم الإحالات من الرؤوس الأضيق إلى الموضوع فى مكان معين إلى تفريع من هذا الموضوع، وذلك على سبيل المثال رأس الموضوع أنهار Rivers (يمكن تفريعها جغرافياً) أى تحت رأس موضوع الأنهار يوجد تفريعات جغرافية عديدة وذلك حتى يمكن أن تؤدى الإحالات إلى أسماء أنهار معينة.

التفريعات الجغرافية وأسماء الأماكن المقسمة بالموضوع:

مشكلة المكان الجغرافى وعلاقته بالموضوع قد تم تناولها بطريقتين مختلفتين فى قائمة رؤوس موضوعات مكتبة الكونجرس، وذلك لأن رؤوس الموضوعات الخاصة بالموضوع يمكن تفريعها بالمكان، كما أن الرؤوس الجغرافية يمكن تفريعها بالموضوع مثال:

Labor Supply - Prance

Ohio - History

ونظراً لعدم وجود قاعدة عامة تشرح الظروف التى تفضل فيها طريقة على أخرى، فمن الأفضل الاعتماد على التعليمات الموجودة تحت رؤوس الموضوعات المخصصة، وذلك لتحديد الطريقة التى يجب استخدامها، فإذا كان رأس الموضوع يحتوى على ملاحظة.

(May Subd geogr)

"يمكن التفريع جغرافيا".

ففى هذه الحالة يتم التفريع الجغرافى من الموضوع، أما إذا لم تكن هناك هذه الملاحظة أو وجد التحديد بين قوسين (Not subd geogr).

(لا يمكن التفريع جغرافيا).

وفى هذه الحالة هناك إحالة عامة إلى تقسيمات فرعية محددة تحت أسماء الأماكن أى أن المساحة الجغرافية المحددة تسبق الموضوع، وذلك على سبيل المثال كما يلى:

History
SA Subdivision History
Under Specific Subjects
And under Name of Countries
States, Cities ... etc.

التاريخ انظر أيضاً

تفريع التاريخ تحت رؤوس الموضوعات المحددة أو تحت اسم البلد أو الولاية أو المدينة ... إلخ.

ويلاحظ أن استخدامات التفريعات تحت أسماء الأماكن يعتبر مشكلة أكبر من المشكلة الخاصة بالتفريعات الجغرافية من الموضوعات، لأن ذلك يستدعى استشارة الدليل للإحاطة بمختلف التفريعات الممكنة، ويلاحظ أن القائمة تستخدم

فرنسا وبريطانيا والولايات المتحدة كنماذج وأمثلة للتعبير عن هذه التفريعات التى يمكن استخدامها، ومع ذلك فيجب أن يوضع التفريع بالتاريخ تحت اسم كل مكان.

التفريعات الحرة أو العائمة Free Floating

يلاحظ أنه حتى عام ١٩٧٤ قام المفهرسون الموضوعيون بوضع رؤوس موضوعات محددة وتوليفات من التفريعات المختلفة وطباعتها فى القائمة، وأما بعد عام ١٩٧٤ فقد تم الاتفاق على أن تكون تفريعات رؤوس الموضوعات فى المستقبل تبعاً لقواعد محددة بدلاً من تخصيصها فى أماكن أو وضع ملاحظات عنها، من أجل ذلك فقد استخدم مصطلح الحرة أو العائمة، ومع ذلك فإن هذه التوليفات لا تظهر كثيراً فى القائمة، أى أن التفريعات التى تظهر هى بقايا من العمل القديم، أو تفريعات مطلوبة لتحديد رأس الموضوع الأضيق، والتفريعات الحرة العائمة يتم تكوينها بواسطة المفهرس فى وقت تحديد أو صياغة رؤوس الموضوعات، بالنسبة لعمل من الأعمال، ومعظم التفريعات الموجودة فى القائمة LCSH موجود بناء على إحالة عامة مثل (Abstracts) أو من خلال إحالة USE العامة، هذا بالإضافة إلى الإرشادات لاستخدام التفريعات العديدة والموجودة ضمن الحواشى، والتى تتبع رأس الموضوع الأساسى فى القائمة، وهناك كشاف هجائى للتفريعات الحرة العائمة مثل Free Floating Subdivision : An Alphabetical index وهذه بالعربية التفريعات العائمة الحرة: كشاف هجائى.

هذا المطبوع هو أداة مفيدة لتحديد التفريعات الحرة العائمة المتوافرة حيث يدلنا على أنواع رؤوس الموضوعات التى يمكن استخدام هذه التفريعات معها، ويجب على المفهرسين الرجوع إلى الدليل لتوليف العناصر بطريقة صحيحة.

رؤوس الموضوعات النموذجية: Pattern Headings

لقد تم فى عام ١٩٧٤ التبنى الرسمى لقاعدة التفريعات الحرة العائمة، والتى يتم التحكم فيها عن طريق الرؤوس النموذجية، حيث وضعت مجموعات معيارية للتفريعات الموضوعية والشكلية لاستخدامها تحت فئات معينة من رؤوس الموضوعات أو الأسماء المستخدمة كموضوعات، وحتى يمكن تجنب تكرار هذه التفريعات تحت جميع الرؤوس الأخرى التى تنتمى لهذه الفئة، ومثل هذه الرؤوس أطلق عليها الرؤوس النموذجية لفئات محددة.

ونظراً لأن العديد من التفريعات التى تعتبر حالياً كتفريعات حرة عائمة تحت رأس موضوع نموذج، هذه التفريعات مطبوعة فى القائمة قبل عام ١٩٧٤ وبالتالى فمازالت تظهر فى أماكن عديدة تحت رؤوس منفردة، ترتبط بفئة معينة. وهناك رؤوس أخرى تتضمن تفريعات يتم التحكم فيها بواسطة الرؤوس النموذجية مازلنا فى حاجة إليها لبناء الإحالات الخاصة بمختلف الرؤوس.

ويلاحظ أن التفريعات العامة الحرة العائمة، لا تتم طباعتها فى القائمة تحت الرؤوس النموذجية، ومع ذلك فهناك بعض التفريعات العامة الموضوعية أو الكلية الحرة العائمة موجودة تحت الرؤوس النموذجية إذا كان ذلك يمثل موضوعا هاما أو شكلا من أشكال المواد داخل الفئة المطلوبة أو أنها مذكورة كأمثلة لإحالات أنظر أيضا SA.

ويلاحظ أن التفريعات التى يتم الانفاق عليها تحت الرؤوس النموذجية، يمكن استخدامها تحت أى رأس موضوع آخر ينتمى لنفس الفئة وبهذه المواصفات، فهى تعتبر تفريعات حرة عائمة، وعلى سبيل المثال فهناك مجموعة من التفريعات التى وضعت لأجزاء الجسم، ومن بين رؤوس الموضوعات المثالية المتصلة بهذه الفئة يمكن أن نجد ما يلى:

Alimentary Canal
Autonomic ganglia
Pituitary body
Renal Artery

وهناك فى القائمة رأسان موضوعيان نموذجيان لأجزاء الجسم وهى مثل:

Foot and heart

والتفريعات الموجودة تحت كل واحد من هذين الرأسين يكون استخدامها كتفريعات حرة عائمة لأى رأس موضوع ينتمى إلى الفئة. وعلى سبيل المثال فإن رأس الموضوع التالى Joints - Biopsy غير مطبوع فى قائمة رؤوس الموضوعات LCSH، ومع ذلك فهو رأس موضوع صحيح نظراً لأن التفريع Biopsy موجود تحت رأس الموضوع Heart. وعلى كل حال فهناك معلومات

إضافية على الرؤوس النموذجية يمكن أن نجدها فى الدليل حيث يبين لنا أشكال الرؤوس الموجودة فى كل فئة من الفئات، فضلا عن قوائم التفريعات التى يمكن استخدامها تحت رؤوس الموضوعات الأخرى، والتى تنتمى لنفس الفئة، وهناك جدول يضم الرؤوس النموذجية، وهذه نفسها مرتبة هجائيا حسب فئات رؤوس الموضوعات وكذلك الرؤوس المحددة التى ستظهر التفريعات تحتها فى قائمة مكتبة الكونجرس، وذلك فى العمود الذى على الجانب الأيمن.

Table of Pattern Headings

Category	*Pattern Heading*
Animals	Fishes
	Cattle
Chemicals	Copper
	Insulin
Colonies	Great Britain—Colonies
Diseases	Cancer
	Tuberculosis
Educational institutions	
Individual	Harvard University
Types	Universities and colleges
Industries	Construction industry
	Retail trade
Languages and groups of languages	English language
	French language
	Romance languages
Legal topics	Labor laws and legislation
Legislative bodies	United States. Congress.
Literary authors (Groups)	Authors, English
Literary works entered under author	Shakespeare, William, 1564-1616. Hamlet.
Literary works entered under title	Beowulf
Literatures (including individual genres)	English literature
Materials	Concrete
	Metals
Military services	United States—Armed Forces
	United States. Air Force.
	United States. Army.
	United States. Marine Corps.
	United States. Navy.
Music compositions	Operas
Musical instruments	Piano
Organs and regions of the body	Heart
	Foot
Plants and crops	Corn
Religious bodies	
Religious and monastic orders	Jesuits
Religions	Buddhism
Christian denominations	Catholic Church
Sacred works	Bible
Vehicles, Land	Automobiles
Wars	World War, 1939-1945
	United States—History—Civil War, 1861-1865

الفصل الثانى عشر

تكشيف واستخلاص الإنتاج الفكرى العربى

دراسة تحليلية

تمهيد :

يكشف دليل "الإنتاج الفكرى العربى فى مجال المكتبات والمعلومات" والذى يحصر هذا الإنتاج على امتداد القرن العشرين كله تقريباً عن عدد قليل جداً من المواد العربية عن التكشيف والاستخلاص، إذ بلغ عدد المواد ١٩٠ مادة منها ٣٨ مادة عن الاستخلاص و ١٤٢ مادة عن التكشيف. وتحاول هذه الدراسة تناول القضايا أو المسائل المتعلقة بالتكشيف والاستخلاص فى العالم العربى وطرق التعامل معها فى ضوء الإنتاج الفكرى العربى فى هذا المجال.

أولاً : بدايات التكشيف :

لعل ما يلفت النظر أن تكون بدايات الكشافات العربية لأعمال مخطوطة وليست مطبوعة، إذ يشير هانى عطية [1] إلى أن كشافات الألفاظ القرآنية قد ظهرت فى منتصف القرن الحادى عشر الهجرى (السابع عشر الميلادى)، حيث ظهر كشاف ترتيب زيبا الذى وضعه الحافظ محمود بن الملادرويش الوردارى (ت ١٠٦١هـ) وهو مقسم على أبواب تتبع حروف المعجم وأورد فيها الآيات ألفبائيا باعتبار الكلمات كما وردت بشكلها الأصلى فى السياق وهى الكلمات المفتاحية مع بيان مواضعها برموز تعرف بها السور. وقد فرغ المصنف من تأليف "ترتيب زيبا" عام ١٠٥٤هـ. ويذكر هانى عطية أيضاً أن الفترة ما بين منتصف العقد الثانى من القرن الثالث عشر الهجرى وحتى أوائل القرن الرابع عشر الهجرى قد شهدت تداخلاً كبيراً بين الكشافات المخطوطة والكشافات المطبوعة فظهر كشاف "نجوم القرأن" لمصطفى بن محمد سعيد الافغان وكان مخطوطاً ثم طبع فى كلكتا عام ١٢٢٥هـ/ ١٨١١م. وقد تتابعت الكشافات القرآنية المطبوعة فى الظهور مع بداية القرن الرابع عشر الهجرى.

---

(١) هانى محيى الدين عطية. كشافات الألفاظ القرآنية المخطوطة: التاريخ والمفهوم. ـ مجلة المكتبات والمعلومات العربية. ـ س ١٧، ع٣ (يوليو ١٩٩٧).ـ ص ٥-٤٢.

فإذا انتقلنا إلى كشافات الكتب العربية المطبوعة نجد أن بداياتها ترجع إلى الثلاثينات من القرن التاسع عشر الميلادى، وأن ظهورها كان "نتيجة لإنتشار حركة الترجمة فى تلك الفترة فى مصر والاحتكاك الثقافى ما بين الثقافة الغربية والثقافة العربية فالكشاف يعتبر سمة أساسية فى الكتاب الأوروبى أخذتها عنه المهاديات المصرية".

وتذكر جيهان السيد أن كتاب "القوانين الداخلية المتعلقة بمشاة عساكر الجهادية" المنشور عام ١٨٣٤ قد ورد فى نهايته "فهرست شامل لخلاصة الخدمات اليومية" وكان عبارة عن قائمة تضم الأفعال التى يقوم بها عساكر الجهادية فى العمود الأول، أما العمود الثانى فبه رتب العساكر المسؤولون عن الأفعال وأمامها أرقام الصفحات التى وردت بها. وهناك كشافات ذات نوعية خاصة تختص بالأبيات الشعرية الواردة فى النص والتى تجمع فى كشاف واحد لتعبر عن شواهد أو دلالات معينة، ومن أمثلة هذا النوع كتاب "شرح شذور الذهب فى معرفة كلام العرب" المنشور سنة ١٨٣٧ إذ ورد الكشاف فى نهايته بعنوان" ما تضمنه هذا الكتاب من الشواهد مرتباً على حروف المعجم"..

وهو عبارة عن أبيات شعرية وردت فى النص مرتبة ترتيبا هجائيا على حروف المعجم وتدل على مواقع معينة لإعرابها مع ذكر أرقام الصفحات.

ومن الكتب المبكرة التى اشتملت على كشافات كتاب "طالع السعادة والإقبال فى علم الولادة وأمراض النساء والأطفال" وهو مترجم إلى العربية ونشر عام ١٨٤٢ وقد ورد فى بدايته بعد قائمة.المحتويات "فهرست كتاب الولادة مرتبة على حروف المعجم" وفيه رتبت الكلمات هجائياً وأقام كل منها رقم الصفحة أو أرقام الصفحات التى ذكرت فيها.

ويشتمل كتاب تاريخ ملوك فرنسا المترجم إلى العربية والمنشور عام ١٨٤٧ على كشاف يقع فى ٦٧ صفحة، ويرد الكشاف بآخر الكتاب وليس أوله كما سبق أن رأينا وهو مرتب ترتيباً هجائياً مع الإشارة إلى أرقام الصفحات وسمى الكشاف "معجم البلدان والأماكن الخفية فى هذا الكتاب". وهكذا كانت الكشافات تقع فى بداية

الكتب، أو فى آخرها وكانت تسمياتها تنحصر فى كلمات مثل "معجم" أو "فهرسة" أو "فهرست" كما كانت تستخدم الترتيب الهجائى [٢] .

أما كشافات الدوريات العربية فقد ظهرت فى أبسط صورها مع ظهور أول الدوريات المصرية وهى La decade Egyptiene (١٧٩٨) فقد كانت هذه الدورية تصدر مرة كل عشرة أيام "وقد صدر لها كشافاً سنوياً يضم ما شملته من مواد خلال العام وكان ذلك أول كشاف سنوى يعد لدورية مصرية ثم ظهرت بعد ذلك قوائم المحتويات السنوية فيما يصدر من مجلات خلال العقود التالية وعلى وجه الخصوص ابتداء من السبعينات من القرن التاسع عشر. ويعتبر هذا النمط هو البذور الأولى لحركة الضبط الببليوجرافى لمحتويات الدوريات.

ومن الكشافات التى صدرت فى الثلاثينات من القرن العشرين "فهرس موضوعات الجرائد العربية اليومية: الأهرام، البلاغ، السياسة، المقطم، قد صدر عام ١٩٣٣، وهناك أيضاً فهرس موضوعات المجلات العربية: الهلال، المقتطف". وقد صدر هو الآخر عام ١٩٣٣.

وقد بدأت الكشافات التركيمية التى تضم مواد إحدى المجلات خلال عدة أعوام فى الصدور منذ أواخر الأربعينات إذ صدر الفهرس العام لمجلة التربية الحديثة الذى يغطى محتوياتها من سنة ١٩٢٧ إلى سنة ١٩٤٧ عام ١٩٤٨.

، كذلك ظهرت الكشافات التجميعية التى تضم تحليلاً لمواد مجموعة من المحلات الصادرة فى مجال موضوعى معين فى نفس الفترة تقريباً (أواخر الأربعينات من القرن العشرين).

وقد تمثل التقدم الحقيقى فى تكشيف محتويات الدوريات العربية فيما ظهر من جهود خلال النصف الثانى من القرن العشرين. حيث أصبحت الكشافات تعد بطريقة

(٢) أ. عايدة إبراهيم نصير. التكشيف المبكر فى العالم العربى ومقترحات للتكشيف الحالى. ــ ص ١٧٩-١٨١ . فــى نــدوة التكشيف والتصنيف فى مراكز المعلومات العربية. ــ الرياض: مكتبة الملك فهد الوطنية، ١٩٩٣.

ب/ جــيهان محمــود الســيد. أوائــل المطــبوعات فــى مصر: دراسة فى الببليوجرافيا التحليلية. ــ الإســكندرية، ٢٠٠٠. ــ ص ٢١٩-٢٢٣ (أطــروحة دكتوراه ــ جامعة الإسكندرية، كلية الآداب، قسم المكتبات والمعلومات)

أكثر تقدماً [٣] .

ثانياً: تكشيف الدوريات والصحف العربية :

تعتبر الدوريات سواء فى شكلها الورقى أو فى شكلها الالكترونى من أهم مصادر المعلومات للباحثين والدارسين، نظراً لاشتمالها على المعلومات الحديثة التى لا توجد فى الكتب عادة، فضلاً عن تناول موادها لموضوعات دقيقة وحيوية. والدوريات من أكثر أوعية المعلومات احتياجاً للكشافات التحليلية التى تظهر محتوياتها، نظرا لتنوع موادها، وامتدادها عبر فترات زمنية طويلة مما يجعل من الصعب إن لم يكن من المستحيل أن يصل الباحث إلى ما يريده منها بسهولة وبسرعة دون وجود كشاف تحليلى. وقد صدر بالعالم العربى نحو عشرة آلاف دورية منها نحو أربعة آلاف دورية جارية على الأقل.

لكن ماذا عن كشافات هذه الدوريات المتوقف منها والجارى؟

من الصعب الإجابة على مثل هذا السؤال نظراً لعدم توافر دراسات عربية حديثة ترصد بصورة كلية أو شمولية وضع الضبط الببليوجرافى لمحتويات الدوريات العربية، ومع هذا سنحاول التعرف على أبرز ملامح الصورة أو الوضع الحالى.

وبداية ليست هناك حاجة لتفصيل القول عن كشافات الدوريات الفردية فهى كثيرة وقد تكون سنوية أو تركيمية (أى تركم محتويات الكشافات السنوية كل بضع سنوات) أو تجميعية (أى تتناول المجلة ابتداء من أول عدد منها حتى أخر عدد متاح عند إعداد الكشاف أو آخر عدد صدر من المجلة.

وعلى سبيل المثال فقد رصدت يسرية زايد فى رسالتها للماجستير (١٩٨٢) [٤] ٧٩ كشافاً لدوريات فردية مصرية، منها مثلاً كشاف مجلة كلية الآداب جامعة

(٣) يسرية محمد عبد الحليم زايد. الضبط الببليوجرافى لمحتويات الدوريات المصرية. ــ القاهرة، ١٩٨٢. ــ ص ٢٢٠-٢٢٢.

(أطروحة ماجستير ــ جامعة القاهرة، كلية الآداب، قسم المكتبات والوثائق)

(٤) المصدر السابق. ــ ص ٣٥٢ - ٣٦٢

القاهرة والذى يغطى الفترة من مايو ١٩٣٣ (تاريخ صدور أول عدد) حتى ديسمبر ١٩٦٣. ومن النماذج الحديثة الكشاف الذى أصدرته دار الكتب والوثائق القومية لمجلة الرسالة. ويقع هذا الكشاف فى خمسة مجلدات وهو يضم نحو ٤٣,٠٠٠ مادة تمثل حياة المجلة ابتداء من عام ١٩٣٣ حتى عام ١٩٥٣. ومن النماذج التى نشرت خارج مصر فهرس المقتطف الذى يغطى محتويات مجلة المقتطف، وهى من أشهر الدوريات فى الوطن العربى وأطولها عمرا، ويغطى الكشاف الفترة من ١٨٧٦ – ١٩٥٢ ويقع الكشاف ــ الذى نشر فى بيروت ــ عام ١٩٦٧ ــ ١٩٦٨ فى ثلاثة مجلدات.

وعموماً فإن هذه الفئة من الكشفات غير منتظمة ولا يُعول عليها كثيرا فى البحث.

أما الفئة الثانية من الكشافات فهى الكشافات التى تحلل محتويات مجموعة من الدوريات العامة التى تغطى موضوعات متنوعة. ولم تحظ هذه الفئة باهتمام كبير، ربما بسبب عدم قيمة الدوريات العربية العامة بالنسبة للبحث العلمى المتخصص، وربما أيضاً بسبب عدم الإقبال على الشراء أو الاشتراك فى مثل هذا النوع من الكشافات من جانب المكتبات مما أدى إلى تخوف الناشرين من تجربة الإصدار المستمر لها، إذ أن التجربتان الوحيدتان فى هذا المجال لم يكتب لهما النجاح أو الاستمرار، فالتجربة الأولى "الكشاف التحليلى للصحف والمجلات العربية" الذى بدأ يصدر على يد مجموعة من المتخرجين فى قسم المكتبات بجامعة القاهرة شهرياً ابتداء من يناير ١٩٦٢ ودون تركيم) ليغطى المحتويات فى أكثر من ثلاثين دورية تصدر فى مصر. وكانت الصحف اليومية فى مقدمة تلك الدوريات المغطاه، إلا أن الكشاف توقف عن الصدور عام ١٩٦٧. [٥] والتجربة الثانية هى "الفهرست" الذى صدر فى بيروت ابتداء من الثمانينات وحتى أوائل التسعينات وكان يغطى مجموعة متنوعة من الدوريات فى مجالات موضوعية مختلفة.

والفئة الثالثة هى فئة كشافات الدوريات الوطنية، أى الكشافات التى تغطى محتويات الدوريات بمختلف أنواعها وموضوعاتها، تلك التى تصدر فى بلد عربى

(٥) محمد فتحى عبد الهادى. التكشيف لأغتراض استرجاع المعلومات. ــ القاهرة: دار غريب للطباعة والنشر والتوزيع، ١٩٨٥ــ ص ١٧٩.

معين. وهذه الفئة هى أحدث الفئات، وهى أكثرها تنظيماً ودقة، وهى أفضلها من حيث الاستمرارية فى الصدور وشمولية التغطية. ومن أبرز النماذج فى هذه الفئة الكشاف التحليلى للصحف والمجلات السورية" الذى يصدر عن مكتبة الأسد الوطنية فى دمشق منذ عام ١٩٨٩، ويغطى محتويات الصحف والمجلات السورية منذ عام ١٩٨٥ حتى الآن.

والنموذج الثانى هو "الكشاف التحليلى للصحف والمجلات السعودية" [٦] الذى بدأ صدوره عن مكتبة الملك فهد الوطنية بالرياض منذ عام ١٩٩٤. ويغطى محتويات أكثر من ٥٠ صحيفة ودورية سعودية.

وقد بدأ الكشاف بمجلد راجع يغطى الدوريات فى الفترة من ١٩٦٦ حتى ١٩٩٢، ثم مجلد يغطى ١٩٩٢ - ١٩٩٣، ويغطى المجلد الثالث ١٩٩٣-١٩٩٤، وهكذا والفئة الرابعة هى فئة كشافات مجموعة من الدوريات المتخصصة فى مجال موضوع معين، وهذه الفئة هى أهم الفئات وهى أكثرها فائدة للباحثين والدارسين فى مختلف المجالات الموضوعية. إلا أنه للأسف الشديد فإن هذه الفئة هى أقل الفئات حظاً بالنسبة للدوريات العربية المتخصصة إذ أنها لم تحظ بالاهتمام الواجب سواء من ناحية الاعداد الفنى أو من ناحية النشر. والنماذج المتاحة قليلة للغاية وتغطى قطاعات موضوعية محدودة. وجدير بالذكر أن الكشاف قد يغطى محتويات الدوريات وحدها أو يضيف إليها بحوث المؤتمرات والكتب وما شابه.

ويعتبر موضوع التربية من الموضوعات التى حظيت باهتمام تكشيفى على المستوى العربى. وتقدم دراسة محمد سالم غنيم عن كشافات الدوريات التربوية المصرية معلومات مهمة عن هذا الموضوع [٧]، فقد أشار إلى أن موضوع التربية فى مصر من أقدم المجالات التى حظيت بجهود التكشيف إن لم يكن أقدمها على الإطلاق، وذكر أن إجمالى عدد الجهود بلغ واحداً وعشرين كشافاً ما بين كشافات

(٦) انظـــر: فؤاد حمد فرسونى. التكشيف الجارى لمحتويات الدوريات السعودية: تحليل الممارسات الناجزة فـــيه وتقويم لدور الكشاف التحليلى للصحف والمجلات السعودية. ــ عالم الكتب. ــ مج ١٦، ع٣ (مايو ــ يونيو ١٩٩٥). ــ ص ٢٣٠ - ٢٦٠.

(٧) محمـــد سالم غنيم. كشافات الدوريات التربوية المصرية: دراسة تحليلية تقويمية. ــ دراسات عربية فى المكتبات وعلم المعلومات. ــ مج ٣، ع٢ (مايو ١٩٩٨). ــ ص ٦٦-٧٨.

فردية [8] وكشافات مجمعة [9] ويهمنا هنا الكشافات المجمعة وهى: الكشاف التربوى (لمركز التوثيق التربوى بالقاهرة)، الببليوجرافية الموضوعية العربية: التربية (المنظمة العربية للتربية والثقافة والعلوم)، دراسة مسحية تقويمية للبحوث التربوية والنفسية (أكاديمية البحث العلمى والتكنولوجيا بالقاهرة)، دليل الدوريات التربوية (المركز القومى للبحوث التربوية والتنمية بالقاهرة) وتبين النتائج التى توصلت إليها الدراسة المذكورة ما يلى:

(أ) غطت بالكشافات الـ ٢١ ست وعشرين دورية بنسبة ٤٩% من المجموع الكلى للدوريات البالغ عددها ثلاثاً وخمسين دورية.

(ب) بلغ مجموع أعمار الدوريات التى تمت تغطيتها فى خدمات التكشيف ٣٤٣ عاماً بنسبة ٥٠,٦٦% من المجموع الكلى لأعمار الدوريات الذى بلغ ٦٧٧ عاما.

(ج) بدا واضحاً قدر من التداخل والتكرار فى التغطية بين خدمات التكشيف المختلفة حيث بلغ مجموع فترات التكرار ٣٠٩ أعوام حوالى ٤٥,٦٤% من مجموع أعمار الدوريات محل اهتمام الدراسة.

(د) بلغ عدد الدوريات التى لم تحظ بالتحليل سبعاً وعشرين دورية بنسبة ٥٠,٩٤% من المجموع الكلى للدوريات التربوية.

ومن الموضوعات الأخرى التى حظيت بتكشيف منتظم نجد مجال المكتبات والمعلومات، إذ قام محمد فتحى عبد الهادى بإعداد دليل "الإنتاج الفكرى العربى فى مجال المكتبات والمعلومات". وهذا الدليل الببليوجرافى يحصر الإنتاج الفكرى العربى فى المجال منذ أواخر القرن التاسع عشر وحتى نهاية عام ١٩٩٦ فى مختلف أشكال أوعية المعلومات، لكن الجزء الأكبر منه هو تحليل للمقالات المنشورة فى الدوريات.

(٨) محمد فتحى عبد الهادى. الإنتاج الفكرى العربى فى مجال المكتبات والمعلومات، ١٩٩١-١٩٩٦م. ــ الرياض: مكتبة الملك فهد الوطنية، ٢٠٠٠.

(٩) محمد فتحى عبد الهادى. التكشيف والاستخلاص/ محمد فتحى عبد الهادى، يسرية محمد عبد الحليم زايد. ــ القاهرة : الدار المصرية اللبنانية، ٢٠٠٠ ــ ص ٦٥-٦٦.

وقد صدر الدليل فى أربعة مجلدات. ويبين المجلد الذى يغطى الفترة من ١٩٩١-١٩٩٦ أن الدوريات المكشفة قد وصل عددها إلى ٢٣٢ دورية [١٠].

ومن الموضوعات التى خظيت بالتغطية على امتداد عدد كبير من السنوات أيضاً "الزراعة" إذ صدر عن المركز القومى للإعلام والتوثيق بالقاهرة "تعريف بالبحوث الزراعية التى أجريت فى مصر فى الفترة من ١٩٠٠ - ١٩٧٠. وقد نشر هذا العمل فى الفترة من ١٩٧٢-١٩٧٨.

تبقى الإشارة إلى جهد كبير نشرته المنظمة العربية للتربية والثقافة والعلوم هو "الببليوجرافيا الموضوعية العربية: علوم الدين الإسلامى". فى عامى ١٩٧٦-١٩٧٧ فى سبعة مجلدات، وهو يغطى المقالات المنشورة فى الدوريات العربية المتخصصة فى علوم الدين الإسلامى والتى بلغ عددها حوالى ١٢٠ دورية عربية صدرت منذ بداية القرن العشرين الميلادى حتى حوالى منتصف السبعينات من هذا القرن.

وهكذا يتضح أن العالم العربى يعانى من نقص واضح فى الكشافات الموضوعية المتخصصة، فإن القطاعات المغطاة محدودة كما سبق أن أشرنا، ونضيف إلى هذا عدم التنسيق والتكامل بين الجهود حتى على مستوى المجال الموضوعى الواحد، هذا فضلاً عن عدم انتظام التغطية وعدم استمرارية الصدور مما يشير إلى أن هناك العديد من البحوث العلمية العربية المدفونة فى الدوريات دون أن يعلم عنها أحداً لغياب المفاتيح الأساسية وهى الكشافات، وتجدر الإشارة إلى أن الجهود السابق الإشارة إلى نماذج منها هى جهود تقليدية، ومعنى ذلك أنه لم يُستفد بعد الاستفادة المطلوبة من التكنولوجيات الحديثة والتى تساعد فى بناء واستخدام قواعد البيانات ذات التغطية الأوسع والأدق والأحدث.

ورغم عدم وجود دراسات عربية حديثة تتناول جهود الاستخلاص فى العالم العربى، إلا أن الصورة على ما يبدو أكثر قتامة من صورة التكشيف. وعموما فمن النماذج فى هذا الصدد "المستخلصات العلمية العربية" التى كانت تصدر عن المركز القومى للإعلام والتوثيق بالقاهرة فى الفترة من ١٩٧٣ - ١٩٧٨، و"المستخلصات

(١٠) عبد الوهاب عبد السلام أبو النور. تكشيف كتب التراث . ــ ص ١٩٧ فى كتابه: بحوث فى المكتبة العربية.ــ الكويت: دار القلم، ١٩٨٥.

الصناعية" التى كانت تصدر عن مركز التنمية الصناعية منذ نوفمبر ١٩٧٢، على أن من أفضل الأعمال الحديثة فى هذا الصدد المستخلصات المصرية العلمية والتكنولوجية التى صدرت عن الشبكة القومية للمعلومات العلمية والتكنولوجية وقد صدر منها عدة مجلدات تغطى العلوم الطبية، والعلوم الهندسية والتكنولوجيا، والزراعة، وهى متاحة فى شكل مطبوع وفى شكل محسَّب أيضاً.

والفئة الخامسة هى فئة كشافات الصحف العربية، وهى فئة متميزة تختلف عن الفئات السابقة بسبب اختلاف الصحف عن الدوريات من نواحى عديدة .. ورغم أن هناك عدة كشافات لصحف عربية بعضها صدر ثم توقف عن الصدور بعد فترة قليلة والبعض الآخر يعد ويستخدم على نطاق ضيق داخل أحد مراكز المعلومات، إلا أن أفضل الجهود العربية فى هذا الصدد هو "كشاف الأهرام" الذى يصدر فى أعداد شهرية عن مركز الأهرام للتنظيم وتكنولوجيا المعلومات منذ يناير ١٩٧٤ مع مجلدات تركيمية، كما قرر المركز إصدار كشافات سنوية على وسائط ميكروفيلمية ثم تركيمية لكل خمس سنوات للأعداد قبل نقطة الابتداء (١٩٧٤) حتى إصدار العدد الأول من الأهرام فى ٥ أغسطس ١٨٧٦. ويستخدم الحاسوب منذ عام ١٩٩٥ فى تبويب وصياغة وتجميع وطباعة الكشاف وبالتالى يستطيع الباحث أن يتعامل مباشرة مع البيانات المختزنة من خلال شاشات الحاسوب، كما يمكنه الحصول على بيان مطبوع بالمادة المطلوبة فى الوقت نفسه. (١١)

ثالثا : تكشيف الكتب العربية :

إن الغرض من إعداد كشاف لكتاب ما هو "أن نبرز كل الألفاظ والمفاهيم ذات الدلالة والتى يشتمل عليها الكتاب بحيث نجعل كل لفظة أو فكرة مدخلا مستقلاً ويسجل أمام المدخل الصفحة أو الصفحات التى ورد بها من النص" (١٢)

---

(١١) المصدر السابق. ص ١٩٨.

(١٢) أ/ التوثـيق ــ كشاف المطبوع (المواصفات القياسية العربية، رقم ٥٢٣-١٩٨٤). ــ المجلة العربية للمعلومات. ــ مج٦، ع١(١٩٨٥). ــ الملحق. ــ ص ٦٠-٦٣.

ب/ توثـيق ــ كشاف المطبوع (المواصفات القياسية المصرية، رقم ١٩١٧-١٩٩٠)ــ القاهرة: الهيئة المصرية العامة للتوحيد القياسى وجودة الإنتاج، ١٩٩٠. ــ ٤ ورقات.

وإذا كان من المألوف أن نجد الكشافات فى الكتب الأجنبية فإنه من المألوف أيضاً ألا نجد الكشافات فى الكتب العربية، فإن نسبة قليلة من الكتب العربية هى التى حظيت بكشافات، وحتى موضوع تكشيف الكتب لم يحظ هو الآخر بدراسات عربية مفيدة ولا تكاد توجد دراسة عربية متعمقة تتناول هذا الموضوع رغم أهميته.

وليس هناك من ينكر قيمة كشافات الكتب وأهميتها فى إبراز محتوياتها على نحو دقيق ومنظم وبما يتيح الوصول إلى المعلومات المطلوبة منها بسرعة وبسهولة ودون جهد أو عناء.

إن الفئة الأولى من الكتب العربية التى حظيت بالتكشيف هى كتب التراث، وربما يرجع ذلك إلى أسباب عديدة منها اهتمام الغربيين فى فترات مبكرة بدراسة كتب التراث العربى الإسلامى ونشرها وتحقيقها، وكان من ضمن عملية التحقيق إعداد فهارس أو كشافات لهذه الكتب لإبراز أهم المعلومات التى اشتمل عليها النص. [١٣] ومن ثم انتقل هذا الفن إلى المحققين العرب والمسلمين. ولعل من أبرز الأسباب التى دعت أيضاً إلى تكشيف كتب التراث هو ضخامة هذه الكتب ووقوع العديد منها فى عدة مجلدات، فضلاً عن اشتمال عدد كبير منها على معلومات موسوعية فى موضوعات شتى، كما أن هذه المعلومات متنوعة ما بين ألفاظ وأماكن وأشخاص وأحاديث نبوية وأشعار، إلخ.

وقد جرت العادة على أن يُعد للكتاب التراثى الواحد عدة كشافات يتناول كل منها جانباً من الجوانب مثل كشاف بالأعلام وآخر بالأماكن وثالث بالقبائل ورابع بالأحاديث النبوية، إلخ. ونادراً ما كان الأمر يقتصر على كشاف قاموسى لكل ما جاء بالكتاب.

---

(١٣) أ/ التوثيق ــ طرق تفحص الوثائق وتعيين موضوعاتها واختيار مصطلحات التكشيف (المواصفات القياسية العربية، رقم ٨٧٨ – ١٩٨٦). ــ المجلة العربية للمعلومات. ــ مج٨ ، ع ٢ (١٩٨٧). ص ١١٧-١٢٦.

ب/ توثيق ــ طرق فحص الوثائق وتحديد موضوعاتها واختيار مصطلحات التكشيف (المواصفات القياسية المصرية، رقم ٢٥٩٥ – ١٩٩٣).ــ القاهرة : الهيئة المصرية العامة للتوحيد القياسى وجودة الإنتاج، ١٩٩٣ . ــ ١١، ١ ورقة.

والفئة الثانية من الكتب العربية التى لقيت اهتماما هى كتب التاريخ والجغرافيا وربما كان ذلك بسبب اشتمالها على أحداث وأماكن متنوعة تستدعى التحليل والتسجيل فى كشافات هجائية.

أما الكتب فى المجالات الأخرى فإنها نادراً ما تشتمل على كشافات بما فى ذلك مجال المكتبات المعلومات نفسه والذى يعتبر التكشيف من أهم موضوعاته.

وجدير بالذكر أن بعض الموسوعات أو دوائر المعارف العربية أو المعربة قد اشتملت على كشافات تحليلية رغم أن المواد بهذه الدوائر أو الموسوعات مرتبة هجائياً. ومن النماذج كشاف "الموسوعة الذهبية" وكشاف "دائرة المعارف الإسلامية".

ومما يلفت النظر أن مُعد الكشاف هو المؤلف نفسه فى أغلب الحالات وكأنه يعتبره جزءاً من عمله فى الكتاب. ورغم معرفة المؤلف الكبيرة بكتابه أفضل من أى شخص آخر إلا أن التكشيف فن وعلم له أصول وقواعد لا يلم بها كل مؤلف. وعموماً فإن ذلك يتم ربما بسبب غياب المؤسسات الببليوجرافية العربية المتخصصة فى التكشيف، وأيضاً غياب المكشفين المتفرغين

تبقى الإشارة إلى نوع خاص من تكشيف الكتب هو تكشيف النصوص، وهو ما سنتناوله بالتفصيل فى نقطة تالية.

رابعاً : أدوات التكشيف والاستخلاص العربية :

يعتمد إنجاز عملية التكشيف على مجموعة من الأدوات الفنية التى تساعد على دقة العمل وتوحيده وتحديثه أو تطويره. ومن أهم الأدوات المكانز، وهى ما سنتاوله فى فصل مستقل، ومن الأدوات أيضاً المواصفات القياسية وأدلة العمل.

وفيما يتعلق بالمواصفات القياسية فإن هناك مواصفتان قياسيتان للتكشيف ومواصفتان قياسيتان للاستخلاص وكلها من إعداد المنظمة الدولية للتقييس وقد قامت المنظمة العربية للمواصفات والمقاييس (قبل إلغائها) بتعريب المواصفات القياسية التى تهم المكتبيين والموثقين العرب تلك التى وضعتها المنظمة الدولية للتقييس (ISO) ومنها المواصفات الخاصة بالتكشيف والاستخلاص.

وقد فعلت نفس الشيء الهيئة المصرية العامة للتوحيد القياسى وجودة الإنتاج حيث قامت بتعريب نفس المواصفات القياسية الخاصة بالتكشيف والاستخلاص واعتبرتها من ضمن المواصفات القياسية الصادرة عنها.

وأول مواصفة هى: التوثيق ــ كشاف المطبوع (المواصفات القياسية العربية، رقم ٥٢٣-١٩٨٤).

ويناظرها فى المواصفات القياسية المصرية: توثيق ــ كشاف المطبوع (رقم ١٩١٧ -١٩٩٠) [١٤]

وتقدم هذه المواصفة مجموعة من القواعد لتمكين الناشرين والمحررين من إعداد كشافات المطبوعات بشكل يسّهل للببليوجرافيين والمكتبيين والباحثين استخدامها. وهى تبدأ بتعريف كشاف المطبوع ثم تتناول أنواع كشافات المطبوعات ومسائل مثل التحرير والرأس العنوان ومعدل الصدور والتصميم الطباعى.

والمواصفة الثانية الخاصة بالتكشيف هى: التوثيق ــ طرق تفحص الوثائق وتعيين موضوعاتها واختيار مصطلحات التكشيف (المواصفات القياسية العربية رقم ٨٧٨ – ١٩٩٦) ويناظرها فى المواصفات القياسية المصرية:

توثيق ــ طرق فحص الوثائق وتحديد موضوعاتها واختيار مصطلحات التكشيف (رقم ٢٥٩٥ – ١٩٩٣) [١٥]

وهذه المواصفة تصف الإجراءات الموصى بها لفحص الوثائق وتحديد موضوعاتها واختيار مصطلحات التكشيف الملائمة ويراد لها أن تكون مرشدا للمكشفين أثناء مراحل تحليل الوثائق وتحديد المفاهيم، كما قد تكون مفيدة لتحليل

(١٤) أ/ التوثـيق ــ المستخلصات للمطبوعات والتوثيق (المواصفات القياسة العربية، رقم ٥٢٥-١٩٨٤). ــ المجلة العربية للمعلومات. ــ مج ٦ ، ع١ (١٩٨٥). ــ الملحق.ــ ص ٨٤-١٠٨.

ب/ توثيق ــ المستخلصات للمطبوعات والتوثيق (المواصفات القياسية المصرية، رقم ٢٦٠١-١٩٩٣). ــ القاهرة: الهيئة المصرية العامة للتوحيد القياسى وجودة الإنتاج، ١٩٩٣.ــ ٨،٣ ورقة.

(١٥) أ / التوثـيق ــ أوراق المستخلصـات فـى المطبوعات المسلسلة (المواصفات القياسية العربية، رقم ٥٨٦-١٩٨٥). ــ المجلة العربية للمعلومات. ــ مج ٧، ع١ (١٩٨٦).ــ ص ٨١-٩٤.

ب/ توثـيق ــ أوراق الاستخلاص فى المطبوعات الدورية (المواصفات القياسية المصرية، رقم ٢٥٩٩ – ١٩٩٣). ــ القاهرة: الهيئة المصرية العامة للتوحيد القياسى وجودة الإنتاج، ١٩٩٣. ــ ٥ ورقات.

استفسارات المستفسرين وترجمتها لأغراض الاسترجاع إلى مصطلحات منضبطة من لغة تكشيف، وكذلك قد تقوم بدور موجه للمستخلصين أثناء إعدادهم للمستخلصات. وتتناول المواصفة عملية التكشيف وغرضها وفحص الوثيقة وتحديد المفاهيم واختيار مصطلحات التكشيف وضبط الجودة.

أما مواصفات الاستخلاص فأولها:

التوثيق ـ المستخلصات للمطبوعات والتوثيق (المواصفات القياسية العربية، رقم ٥٢٥-١٩٨٤).

ويناظرها فى المواصفات القياسية المصرية :

توثيق ـ المستخلصات للمطبوعات والتوثيق (رقم ٢٦٠١ – ١٩٩٣) [١٦]

وتختص هذه المواصفة بتقديم إرشادات لإعداد وعرض مستخلصات الوثائق، وهى تركز على المستخلصات التى يعدها مؤلفو الوثائق الأولية وعلى نشرها، لأنه يمكن لهذه المستخلصات أن تكون مفيدة لقراء هذه الوثائق إلى جانب إمكانية الاستنساخ فى المطبوعات والخدمات الثانوية كما هى وبدون أى تعديل أو بإدخال تغييرات طفيفة عليها، وهى تناسب أيضاً قيام الآخرين من غير المؤلفين بإعداد المستخلصات ومن ثم تضمنت أيضاً إرشادات خاصة بعرض المستخلصات فى المطبوعات والخدمات الثانوية. وتتناول المواصفة الغرض من المستخلصات واستخداماتها ومعالجة محتويات الوثيقة والعرض والأسلوب، ويلحق بالمواصفة أمثلة لمستخلصات.

والمواصفة الثانية الخاصة الاستخلاص هى:

التوثيق ـ أوراق المستخلصات فى المطبوعات المسلسلة (المواصفات القياسية العربية، رقم ٥٨٦-١٩٨٥)

ويناظرها فى المواصفات القياسية المصرية:

توثيق ـ أوراق الاستخلاص فى المطبوعات الدورية (رقم ٢٥٩٩ – ١٩٩٣) [١٧]

(١٦) جامعة الدول العربية. الأمانة العامة. مركز التوثيق والمعلومات. الدليل العملى للتحليل الموضوعى والتكشيف/ اشراف محمود أحمد إتيم. ـ القاهرة: المركز، ١٩٨٧. ـ ٢٢٣ص.

(١٧) على السليمان الصوينع. استرجاع المعلومات فى اللغة العربية. ـ الرياض: مكتبة الملك فهد الوطنية، ١٩٩٤. ـ ١٧٦ ص

وورقة الاستخلاص هى صفحة، يفضل أن تكون منفصلة، توضع فى بداية أو نهاية كل دورية أو أى مطبوع مسلسل آخر وتشتمل على أكثر من إسهامة واحدة وتتضمن وصفاً ببليوجرافيا ومستخلصا لكل إسهامه.

وتختص هذه المواصفة القياسية بوضع قواعد تقديم ورقة الاستخلاص فى الدورية أو أى مطبوع مسلسل آخر، وهى تتناول نص ورقة الاستخلاص والتقديم المادى ولغة أو لغات ورقة الاستخلاص، ويلحق بها نموذج لورقة استخلاص بالعربية وآخر لورقة استخلاص باللغة الإنجليزية.

والملاحظ بصفة عامة التكرار الواضح بين المواصفات القياسية العربية والمواصفات القياسية المصرية، والاختلافات هى فقط فى الألفاظ التى يستخدمها كل مترجم أو معرب، وفى الأمثلة فى الملاحق والمصطلحات.

كما يلاحظ عدم تطوير أو تحديث هذه المواصفات القياسية منذ تعريبها لأول مرة.

وفيما يتعلق بأدلة العمل فإننا نلاحظ نقصا واضحاً فى أدلة العمل العربية رغم أهمية مثل هذه الأدلة فى إتقان العمل وجودته وفى تطويره وتحديثه وأيضاً فى تدريب الموظفين الجدد. ولا نصادف من الأدلة المنشورة سوى دليل واحد صدر عن مركز التوثيق والمعلومات بجامعة الدول العربية عام ١٩٨٧ بعنوان "الدليل العملى للتحليل الموضوعى والتكشيف" (١٨).

وهذا الدليل أعد فى الأساس لإرشاد المكشفين فى مركز التوثيق والمعلومات بجامعة الدول العربية، ولكن يمكن أن يستفيد منه المكشفون فى المكتبات ومراكز المعلومات العربية الأخرى. وذكر فى مقدمة الدليل أنه روعى فى إعداده أن يكون على درجة عالية من الدقة والتركيز والبساطة كما أنه أعد من واقع الخبرة والتجربة العملية التى مر بها المركز عبر سنوات تأسيسه. ويتناول الدليل العلاقة بين التصنيف والتكشيف والوصف الببليوجرافى أو الفهرسة.، كما يتناول الدليل المكنز باعتباره الأداة الرئيسة المستخدمة فى عملية التكشيف، وخطوات تحليل المحتوى

---

(١٨) ناصر محمد السويدان. الاسترجاع الموضوعى بواسطة كلمات العنوان. ــ ص ٥٣٣-٥٦٨. فى: السجل العلمى لندوة استخدام اللغة العربية فى تقنية المعلومات. ــ الرياض: مكتبة الملك عبد العزيز العامة، ١٩٩٣.

الموضوعى، ثم يتناول تكشيف الأنواع الخاصة من المواد وهى المقالات واختيارها من الدوريات وتكشيفها، وتكشيف وثائق جامعة الدول العربية. وينتهى الدليل بعدد من الملاحق منها: مواصفة عرض الوصف الببليوجرافى وقواعد النقحرة ومستويات التكشيف للوثائق المختلفة ونظام تصنيف وثائق جامعة الدول العربية.

خامساً : تكشيف الكلمات الدالة فى السياق:

إن استخدام الكلمات المفتاحية أو المهمة فى عناوين الوثائق كمداخل موضوعية وترتيبها هجائياً بعدد مرات ورودها مع الإشارة إلى السياق الذى وردت فيه ليس أمرا جديداً فهو مطروح منذ أواخر الخمسينات من القرن العشرين الميلادى فى الإنتاج الفكرى الغربى، وقد أدخلت عليه بعد ذلك تحسينات وتعديلات كثيرة، كما استخدم الحاسب الإلكترونى فى إنتاج كشافات كثيرة من هذا النوع، لكن النظام لم يلق قبولا على نطاق واسع بالنسبة للإنتاج الفكرى العربى، فالكتابات عنه قليلة أو محدودة وإعداد كشافات الكلمات الدالة فى السياق لعناوين الوثائق العربية لم يتعدى مرحلة التجريب على حد علم الباحث، وإن كان شائع الاستخدام بالنسبة لكشافات التباديل لبعض المكانز العربية (انظر مكنز الفيصل مثلاً). وليس هنا مكان عرض كيفية إعداد هذا النوع من الكشافات لكننا نشير إلى أن من أفضل الدراسات العربية عن هذا الموضوع دراسة على السليمان الصوينع بعنوان "استرجاع المعلومات فى اللغة العربية". (١٩)

يشير الصوينع إلى أنه رغم قيمة اللغة الطبيعية فى الاسترجاع إلا أن هناك مشكلات دلالية وتركيبية واسترجاعية تعيق هذا الاسترجاع فى اللغة العربية منها:

١- التصريف والاشتقاق، حيث تمتاز اللغة العربية بالثراء فى التصريف والاشتقاق مما يشتت المداخل المتشابهة ويعيق البحث والاسترجاع فإذا كان مصطلح البحث كلمة (إدارة) مثلاً فإن لهذه الكلمة تصاريف وصيغاً إعرابية

---

(١٩) أ/ عـبد الرحمن فراج. تحليل الاستشهادات المرجعية: بعض مشكلاته فى الإنتاج الفكرى العربى:ـــ مجلة المكتبات والمعلومات العربية. ـــ س. ١ ، ع١ (يناير ١٩٩٠). ـــ ص ٧٩-١٠٣

ب/ عبد الرحمن فراج. تكشيف الاستشهادات المرجعية. ـــ مجلة المكتبات والمعلومات العربية . ـــ س ١١ ، ع٢ ، ٣ (إبريل/ يوليو ١٩٩١). ـــ ص ١٣٦-١٧٦.

متعددة يؤثر تفاوتها على البحث والاسترجاع مثل إدارات، إدارى، مدير، مدراء، يدير، تدير ... وبعض هذه التصاريف قد تلحقها زوائد متعددة مثل حروف الجر المتصلة وآل التعريف.

ولعل مشكلة تشتت المداخل واختلاف ترتيب الكلمات الدالة على مفهوم واحد تتضح فى التفاوت بين شكل الكلمات فى صيغتى المفرد والجمع مما يستوجب البحث تحت مصطلحين دالين على مفهوم واحد (طفل، أطفال)

٢- المترادفات، حيث تكثر فى اللغة العربية الكلمات المترادفة الدالة على مفهوم واحد.

٣- الألفاظ المشتركة أو المتجانسة، حيث تأخذ الكلمة الواحدة عدة معانى ومفاهم.

٤- تغير دلالة المفاهيم المترابطة حسب التركيب وترتيب المفاهيم أو الموقع الإعرابى سواء كانت الكلمة الدالة مضافاً أو مضافاً إليه.

٥- قد تأتى المفاهيم على صيغ أفعال مثل: كيف تحج أيها المسلم؟

٦- الإملاء ورسم الكلمات.

٧- استخدام المجاز والرمز أو استخدام مفاهيم لا علاقة لها بالنص أو استخدام عناوين غريبة أو مبهمة خاصة فى الموضوعات الأدبية والفنية.

ويمكن أن نضيف إلى ما سبق صعوبة تحديد كلمات التوقف العربية بشكل دقيق وقاطع، إذ عند تتبع العناوين العربية نجد أنها قلما تخلو من حروف أو مفردات حشو يستخدمها المؤلفون باللغة الطبيعية مثل "حول" الثقافة العربية. ويتطلب الأمر حصر كلمات التوقف الممكنة بالاعتماد على دراسة إحصائية ولغوية لمعرفة أنواع ودلالات الكلمات فى عناوين المطبوعات العربية.

وعلى العموم فإن مناقشة الصوينع لمشكلات هذا النوع من التكشيف والاقتراحات التى قدمها جديرة بالنظر والاهتمام.

فإذا انتقلنا إلى التجارب العربية فى هذا المجال فإن من أهمها ما قام به ناصر السويدان [٢٠] فى تجربة عملية تهدف إلى التعرف على مدى كفاية كلمات العنوان

---

(٢٠) أحمد محمد الشامى. المعجم الموسوعى لمصطلحات المكتبات والمعلومات: إنكليزى ـ عربى/ أحمد محمد الشامى، سيد حسب الله. ـ الرياض: در المريخ للنشر، ١٩٨٨. ـ ص ٢٩٩.

للدلالة على الموضوعات فيما ينشر فى الدوريات العربية. وقد اعتمدت الدراسة على عينة من ثمان عشرة دورية من تخصصات مختلفة سجلت حاسوبياً وتم تحليل بياناتها واستخلاص أهم النتائج. وقد بلغ عدد العناوين التى تم تسجيلها ١٧٨ عنواناً، ونتيجة لمعالجة البيانات حاسوبياً فقد بلغ عدد العناوين ١٠١٢ عولجت وفقاً لمبدأ التباديل أى تكرار عرض العناوين بحسب عدد كلماتها الدالة.

وقد أظهر البحث أنه "يمكن الاستفادة من كلمات العنوان للاستدلال على الموضوعات المنشورة فى الدوريات العربية، إلا أن تمثيل العنوان لموضوع الوثيقة يتفاوت من عنوان إلى آخر، فبعضها لا تدل الكلمات فيه على الموضوع نهائياً وبعضها يشير من بعيد إلى الموضوع بدون تحديد وفئة ثالثة من العناوين نجد فيها كلمة واحدة تدل على الموضوع بينما عناوين أخرى تدل فيها أكثر من كلمة على العنوان أى أن تمثيل العناوين للموضوعات هو تمثيل جزئى. ومن خلال استعراض حرف الألف من الكشاف ظهر أن الثلث تقريباً من الكلمات المفتاحية تدل على الموضوع بينما نسبة مماثلة تقريباً لا تدل من قريب أو بعيد على الموضوع ويبقى بين هذا وذاك نسبة الثلث تشير من بعيد أو قريب إلى الموضوع".

وقد أظهر البحث أيضاً العديد من الصعوبات، منها ما يعود إلى المؤلفين فى اختيارهم لعناوين ذات كلمات غير دالة لأسباب مختلفة، أو لطبيعة الموضوع مثل الشعر والقصة، أو لطبيعة اللغة مثل التصاق حروف الجر بالكلمة، أو لأسباب ترجع إلى الحواسيب".

وقد رأى الباحث أن التكشيف بواسطة كلمات العنوان ليس بديلا كافياً عن التكشيف المقيد، وأنه عند تطبيق هذا النوع من التكشيف بواسطة كلمات العنوان ينبغى وضع ضوابط واحتياطات من شأنها تحسين مستوى الأداء، وأنه من المناسب أن يقتصر استخدام الكشاف على البحث الآلى المباشر فى قاعدة البيانات ولا ينصح بطبع الكشاف بسبب كبر حجمه إذا وضع على الورق.

سادساً : تكشيف الاستشهادات المرجعية فى المصادر العربية :

لعله مما يدعو للأسف أنه لا توجد حتى الآن أية كشافات عربية للاستشهادات المرجعية رغم قيمة هذا النوع من الكشافات فى التعرف على خصائص الإنتاج

الفكرى فى موضوع ما. وفى هذا الصدد يذكر عبد الرحمن فراج [٢١] أن عبء إعداد وإنشاء كشافات الاستشهاد المرجعى العربى. ويكاد يقع على كاهل الدارسين الذين يتعاملون أساساً مع تحليل الاستشهاد المرجعى كأحد أساليب منهج القياسات الببليوجرافية وذلك فى سياق إعدادهم للدرجات الأكاديمية المختلفة.

والحقيقة أن هناك دراسات عربية عديدة، أغلبها رسائل للماجستير والدكتوراه، تتناول خصائص الإنتاج الفكرى العربى فى موضوعات عديدة، وتعتمد فى ذلك على منهج القياسات الببليوجرافية بما فى ذلك استخدام أسلوب تحليل الاستشهادات المرجعية اعتماداً على الجهد الذاتى وليس على كشافات منشورة.

وتتعرض دراسة عبد الرحمن فراج السابق الإشارة إليها (فى جزئها الأول) للمشكلات التى يواجهها محلل الاستشهادات المرجعية عند فحصه للإنتاج الفكرى العربى فى سياق تكشيف هذا الإنتاج وتحليله. وتدور هذه المشكلات حول تعدد مواقع الإشارات الببليوجرافية واختلاف طرق عرضها أو تسجيلها ومدى اكتمالها فى البحوث العربية. وفى الجزء الثانى من دراسته يتناول فراج الخطوات الرئيسية لإعداد وإنشاء كشاف الاستشهاد المرجعى بشكل يدوى وينتهى بنماذج لبعض أنماط الاستشهاد المرجعى من واقع ببليوجرافيات وهوامش بعض المؤلفات العربية.

وعلى العموم فإنه يبدو من الضرورى أن تضطلع إحدى الهيئات العربية بإعداد مثل هذا النوع من الكشافات لما له من فائدة كبيرة فى البحث العلمى.

سابعاً : تكشيف النصوص العربية بلغتها الطبيعية :

كشاف النص هو "كشاف هجائى لجميع الألفاظ أو لأكثر الألفاظ أهمية فى أى عمل مع ذكر السياق الذى وردت فيه وبيان موضعها" [٢٢]. وعادة ما تعد كشافات النصوص للكتب المقدسة والقوانين والتشريعات والأعمال الأدبية لكبار الأدباء.

---

(٢١) علـى السـليمان الصـوينع. كشـافات النصوص وتطبيقاتها فى نصوص القرآن والحديث. ــ مجلة المكتبات والمعلومات العربية. ــ س ٧، ع٣ (يوليو ١٩٨٧). ــ ص ١٢.

(٢٢) سعود عبد الله الحزيمى . ــ دليل المراجع/ إعداد سعود عبد الله الحزيمى، بسام عبد الغنى صبرة ــ الرياض: مكتبة الملك فهد الوطنية، ١٩٩٤. ــ ص ٣١١ ــ ٣١٤.

وتعد كشافات النصوص العربية من أهم الأنماط التى اهتم بها العرب والمسلمين وخاصة بالنسبة للقرآن الكريم والحديث النبوى.

ويذكر الصوينع [٢٣] أن الكشافات الخاصة بنصوص القرآن الكريم والحديث النبوى لها بعض الفوائد مثل :

(أ) للاستشهاد بنصوص القرآن الكريم أو الحديث النبوى لأغراض مختلفة.

(ب) التأكد من دقة نص الآية أو الحديث المنقول.

(ج) التأكد من مكان الآيات القرآنية فى السور.

(د) معرفة مكان نزول الآيات سواء كانت مكية أو مدنية.

(هـ) معرفة عدد مرات ورود بعض الألفاظ فى القرآن الكريم واستخداماتها المتعددة.

(د) التأكد من صحة الأحاديث النبوية ومعرفة أسانيدها.

ويورد "دليل المراجع" الصادر عام ١٩٩٤ [٢٤] بيانات عن ٢٣ كشافاً للقرآن الكريم.

وقد شهدت برامج القرآن الكريم الآلية تطوراً متزايداً منذ عام ١٩٧٨ حتى الآن. ويذكر هانى عطية [٢٥] أن برنامج القرآن الكريم"العالمية" وبرنامج القرآن الكريم "باك" هما من أفضل البرامج، فكلاهما يعتمد الرسم العثمانى واستراتيجية بحث عالية.

وعموماً فقد بدأت كشافات القرآن العربية والإسلامية فى الصدور منذ أواخر القرن التاسع عشر الميلادى فى مختلف البلاد الإسلامية مثل تركيا وإيران والهند ومصر. ولعل أقدمها هو كتاب "دليل الحيران فى الكشف عن آى القرآن" تأليف صالح ناظم بن محمد، وقد نشر فى الآستانة سنة ١٢٨٤هـ.

(٢٣) هانىء محيى الدين عطية. برامج القرآن الآلية: دراسة نقدية . ـ مجلة المكتبات والمعلومات العربية. ـ س ١٤، ع٣، ٤ (يوليو ـ أكتوبر ١٩٩٤). ـ ص ٢٨

(٢٤) سعود عبد الله الحزيمى. كشافات النصوص العربية. ـ مجلة مكتبة الملك فهد الوطنية. ـ مج ٢، ع ١ (يونيو ـ ديمسبر ١٩٩٦). ـ ص ١٢٣ .

(٢٥) مساعد بن صالح الطيار. كفاءة التحليل الصرفى فى استرجاع النصوص العربية. ـ مجلة، مكتبة الملك فهو الوطنية. ـ مج٤، ع١. (مايو ـ أكتوبر ١٩٩٨). ص ٧-٢٣.

ومن أشهرها "المعجم المفهرس لألفاظ القرآن الكريم" الذى أعده محمد فؤاد عبد الباقى ونشر بالقاهرة عام ١٣٦٤هـ.

وإضافة إلى كشافات القرآن الكريم فإن هناك كشافات للحديث النبوى، وهناك عدة مناهج أتبعت فى معالجة الأحاديث يهمنا منها هنا المنهج الحديث وهو تكشيف النصوص كاملة، وقد أخذ بهذا المنهج المستشرق فنسنك فى الكتاب الذى أشرف على إعداده وهو "المعجم المفهرس لألفاظ الحديث النبوى" وقد صدر فى سبعة مجلدات على مراحل ابتداء من عام ١٩٣٦ حتى عام ١٩٦٩، وفى عام ١٩٨٨ صدر المجلد الثامن بعنوان الفهارس.

ونشير إلى أنه رغم كثرة كشافات القرآن الكريم والحديث النبوى، إلا أن هناك صعوبات فى استخدام بعض هذه الكشافات بسبب عدة عوامل منها الترتيب حسب أصول الكلمات. ومع هذا فإن استخدام النظم الآلية الحديثة فى استرجاع النصوص العربية فى هذا المجال يعمل على تحسين الوصول إلى المطلوب بسهولة وبسرعة وقد تقدمت هذه النظم تقدماً ملحوظاً فى السنوات الأخيرة. ومن أهم الدراسات فى هذا الصدد دراسة مساعد الطيار. فقد أشار إلى أن موسوعة الحديث النبوى (من إنتاج شركة العالمية للبرامج) من أهم أنظمة استرجاع النصوص التى استخدمت تقنية التحليل الصرفى فى استرجاع النصوص العربية. وقد قام الباحث باستخدام هذه الموسوعة فى إجراء عشر استفسارات من أجل مقارنة أداء ثلاثة مناهج من مناهج البحث فى أنظمة الاسترجاع العربى وهى: البحث بمستوى الكلمة، والبحث بمستوى الكلمة مع اللواصق، والبحث بمستوى جذر الكلمة. وهدفت الدراسة إلى تسليط الضوء على أداء هذه المناهج الثلاثة مع استخدام قياس التحقيق الاستدعاء، أو تسليط الضوء على أثر التحليل الصرفى فى أداء استرجاع النصوص العربية.

وقد أظهرت هذه الدراسة التجريبية أن البحث بمستوى الجذر يسترجع نصوصاً أكثر من المنهجين الآخرين (كلمة، كلمة مع اللواصق)، ولعل هذا يعود إلى قدرة البحث بمستوى الجذر على استرجاع جميع الصيغ للاستفسار والتى تشترك فى الجذر نفسه. وقد استرجع البحث بمستوى الجذر ٨١% من النصوص الصالحة، بينما استرجع البحث بمستوى الكلمة نسبة ١٨% فقط من النصوص الصالحة واسترجاع البحث بالكلمة مع اللواصق نسبة ٢٩% من النصوص

الصالحة. وفيما يتعلق بأداء التحقيق فقد تبين أن البحث بمستوى الكلمة حقق أعلى نسبة تحقيق مقارنة بالمنهجين الآخرين عند مستوى ٩٦% وحقق البحث بمستوى الكلمة مع اللواصق نسبة ٩٤% وكانت أدنى نسبة تحقيق هى من نصيب البحث بمستوى الجذر وذلك عند مستوى ٧٥%. ويخلص الباحث إلى أن البحث بمستوى الجذر يرفع من أداء النظام خصوصاً فيما يتعلق بالاستدعاء (جميع الوثائق أو النصوص ذات العلاقة أو القريبة من موضوع البحث)، كما يرى الباحث أداء التحقيق (الوثائق أو النصوص ذات العلاقة القوية بالموضوع) بالنسبة للبحث بمستوى الجذر يمكن تحسينه إذا كان هناك نوع من العمق فى التحليل اللغوى للجذور العربية. وهو يرى كذلك أن استخدام البحث بمستوى الجذر مهما فى بعض المجالات مثل البحوث فى النصوص الشرعية أو القانونية بينما قد لا يكون مناسباً للنصوص العامة أو الشاملة.

ونحن مع الباحث فى أن قضية أثر اللغة فى استرجاع المعلومات باللغة العربية لازالت بحاجة إلى مزيد من البحث والدراسة.


Converted by Tiff Combine - (no stamps are applied by registered version)

الفصل الثالث عشر

أضواء على التطورات المعاصرة ومستقبل خدمات التكشيف والاستخلاص في القرن الحادي والعشرين

مقدمة

يبدو أن بحوث التكشيف والاستخلاص فى تناقض مستمر خلال السنوات الأخيرة، على الرغم من الحاجة المتزايدة لتحسين نوعية التكشيف والاستخلاص، آخذين فى الاعتبار فاعليه التكلفة.. ولقد تركزت معظم جهود تطوير البحوث فى المجال، على كيفية استبدال المصادر المحسية بالمصادر الإنسانية، نظرا للفيض الهائل من المعلومات المنتجة كل يوم، أى أن التركز هو على استخدام النظم الآلية وليس على التعرف على احتياجات المستفيدين ــ بمستوياتهم المعرفية المختلفة ــ من الكشافات والمستخلصات.

وتتناول هذه الدراسة فى البداية خريطة بحوث التكشيف والاستخلاص شاملة للجوانب المعرفية والتحكم فى المصطلحات والتكشيف والاستخلاص بمعاونة الحاسب، وبحوث عرض وتركيب الكشافات، ثم تقييم الكشافات والمستخلصات الآلية ثم تتناول الدراسة نماذج من أهم كشافات الوثائق المنتجة آليا والتى ينتظر لها الذيوع والانتشار فى القرن الحادى والعشرين، ثم يقوم الباحثان بتحليل نماذج من الانتاج الفكرى الحديث (١٩٩٧- ٢٠٠٠م) للتعرف على اتجاهاته ولغاته، واهتمامه بالتفاعل مع التخصصات الأخرى خصوصا الحاسبات واللغويات والذكاء الاصطناعى والنظم الخبيرة وبعض الاتجاهات العامة فى النظرية والممارسة وأخيرا النتائج والتوصيات اللازمة لدعم وتطوير البحوث فى المجال باللغة العربية.

أولا:خريطة بحوث التكشيف والاستخلاص

يركز هذا الجزء على التكشيف باعتباره الخطوة الأولى لإعداد المستخلصات، وإن كان الباحثان سيتناولان الاستخلاص بالتحديد فى بعض المواقع، ويجب الإشارة منذ البداية أنه وعلى، غير المتوقع فإن عدد بحوث التكشيف لا تتزايد على مرور السنين، ولكنها تقل، وقد أكدت هذا الاتجاه الباحثة جيسكا مايلستد (Milstead, J.,

1994p. 5774) فقد جاء فى دراستها أن مجلة المراجعة السنوية لعلم وتكنولوجيا المعلومات(ARIST) فى سنواتها العشر الأولى (١٩٦٦- ١٩٧٥) قد نشرت فى المتوسط (٢٥- ٣٠) مدخل كشفى كل عام، تحت رأس موضوع التكشيف والإحالات الموضوعية، ثم انخفض هذا المتوسط السنوى إلى (١٠- ١٢) مدخل كل عام، وإذا كان من العسير وضع التبرير الحقيقى لهذا الانخفاض التدريجى للبحوث، فهناك من يذهب إلى أن معظم بحوث ودراسات التكشيف والاستخلاص نظرية، وليس فيها نظم تطبيقات فعلية تستجيب لاحتياجات الحياة اليومية إلا أقل القليل، أما خريطة بحوث التكشيف المطلوبة من الناحيتين النظرية والعملية فتتضمن ما يلى:

١- الجوانب المعرفية Cognitive aspects

وهذه الجوانب تتضمن

أ. عمليات معرفية فى التكشيف: أى التعرف على ما تقوله الوثيقة aboutness والأسئلة التى يمكن أن تجيب عليها، وترجمة ذلك فى مصطلحات يستخدمها نظام التكشيف سواء كان النص الحر أو المصطلحات المحكومة.. والفجوة هنا هى فى عدم امكانية الحاسب فى انتاج كشافات تضاهى ما ينتجه الإنسان.. فضلا عن اختلاف اثنين من الناس عند قيامهم بتكشيف نفس الوثيقة نظر لاختلافهم فى مستوى فهم الموضوع أو عدم استيعابهم لدلالات المصطلحات أو غير ذلك وهذا اختلاف فى الممارسة المعرفية.

ب. عمليات معرفية فى البحث: أى العمليات التى تخص المستفيدين من الكشاف أى مدى الإفادة Usability، كيف يقوم الباحثون بصياغة اسئلتهم، كيف تساعد الكشافات فى تحديد احتياجاتهم الفعلية؟ إن تنظيم الكشافات لتتلاءم مع العمليات المعرفية للمستفيدين يتناول قضايا الإخراج وتركيب النص Structure وترتيبه وتجميع المفاهيم المتقاربة مع بعضها.

٢- التحكم فى المصطلحات Vocabulary Control

أ . علاقات المكنز:

هل يعرف المستفيدون معانى مختلف هذه العلاقات (BT/NT/RT, USE/UF) .....؟. ولا نستطيع القول فى الوقت الحاضر رغم كل تطورات استخدام اللغة الطبيعية أننا نستطيع أن نستبعد المصطلحات المحكومة.. أى أن الوقت مازال بعيدا

قبل أن نجزم بأن محركات بحث استرجاع النص الكامل ستكون من القوة بحيث تستبعد تماما التكشيف الإنسانى.

ب. **حجم الكنز:** لقد وصلت بعض المكانز إلى حجم ضخم كالمكنز الهندسى Engineering Thesaurus (١٦٠,٠٠٠ مصطلح) فما الحجم المناسب للمكنز فى علاقته بقاعدة البيانات؟ وهل هناك توازن يمكن الوصول إليه بين استخدام المكنز ودقة الاسترجاع؟ وعند أى نقطة سيصبح المكنز ضخما للدرجة التى تكون فيها الوثائق ذات الموضوع الواحد مشتتة تحت مصطلحات مختلفة.

ج. التوازن بين التحكم فى المصطلحات واللغة الطبيعية:

البحث هنا عن التوازن بالنسبة للكشاف الالكترونى، خصوصاً وهناك من الدراسات التى تشير إلى أن أفضل استرجاع هو مزيج من التكشيف الإنسانى باستخدام المصطلحات المحكومة والقدرة على بحث النص الكامل، ومع ذلك فالاتجاه المعاصر يتجه نحو زيادة قدرات برامج الاسترجاع.

٣- التكشيف والاسترجاع بمعاونة الآلة

Machine aided Indexing and Retrieval

أ. **نظم المصطلحات المقترحة:** Candidate Term Systems

يستخدم العديد من منتجى قواعد البيانات برامج التكشيف التى تستعين بالحاسب، وهذه البرامج تقوم بتحليل النص (عادة المستخلصات) ثم تقترح مصطلحات ممثلة لموضوع النص، وهنا تقع المسئولية على المكشف لمراجعة هذه المصطلحات المقترحة يحذف منها أو يضيف إليها ما يراه، ومن أمثلة هذه النظم المطبقة فعلا ما هو موجود فى معهد البترول الأمريكى وفى مركز المعلومات الفنى الوطنى وفى ناسا (NASA)... وهناك بعض الشركات الخاصة التى قامت مؤخرا بتقديم خدماتها فى هذا التكشيف الذى تعاونه الآلة.. ولكن ما هو أثر هذا الاتجاه على نوعية التكشيف، وإلى أى مدى يقترح النظام المصطلحات المفيدة التى ربما فاتت على المكشف اليدوى؟ وما هى تكاليف هذه النظم على المدى البعيد؟.

ب. استكمال التكشيف الإنسانى Supplementing Human Indexing:

هناك اتجاه آخر لا يبدأ من الآلة بل من الإنسان حيث يقوم المكشف بإعداد مدخلاته من المصطلحات، التى يتم اقتراح مصطلحات إضافية عن طريق الحاسب، وتعتبر المكتبة الوطنية واحدة من أهم المؤسسات التى تطبق هذا النظام والمسمى (Med In Ex) والمهم أن فاعلية التكلفة الخاصة بهذه النظم غير معروفة تماما (Mistead, J., 1994, P 580).

ج. الجمع بين التكشيف اليدوى واسترجاع اللغة الطبيعية:

هناك نظم فى الوقت الحاضر تجمع بين التكشيف الإنسانى والأساليب الإحصائية والدلالية Semantic والنحوية Syntactic للاسترجاع، آخذه فى اعتبارها فاعلية التكلفة، وواضح أن هذه النظم تعمل اعتمادا على بعض الجهد الإنسانى فى التحليل، ولكن ما هى الأجزاء التى يمكن أن يحل محلها استرجاع اللغة الطبيعية حتى يمكن الوصول إلى نتائج تعاونية ناجحة؟ ما زال الأمر قيد الدراسة والبحث.

د. نظم الفئات Categorization System

هناك دراسات جارية تتعلق بتنظيم النصوص إلى مجموعة فئات Categories محددة سلفا، ويذهب القائمون على هذه النظم بإمكانية الوثوق بها (مثل المتبع فى النظام الدولى للمعلومات الذرية) وهذا النظام هو جزء من عملية ضبط الجودة حيث يستخدم مجموعة من القواعد لضبط الفئات مع مصطلحات التكشيف، وإذا ما حدث خلل فى تشغيل النظام فيرجع إلى المكشف الإنسانى.

٤- العرض والتركيب Layout and Structure

تنسحب المناقشة هنا على الكشافات المطبوعة والالكترونية كما أن الأمثلة التى ستعرض هنا هى مجرد نماذج وليست حصرا للمجال.

أ. الاحالات Cross- references

هل يدرك المستفيدون حقا المقصود بإحالة "انظر"، انظر أيضاً والفرق بينهما؟ وما هو أقصى عدد من هذه الاحالات يكون مفيدا؟ أى الحد الذى سيقف عنده الباحث فى الاستعانة بهذه الأدوات؟ خصوصا وقد أظهرت بعض البحوث الارتباك والاحباط

الذى اصاب بعض المستفيدين من جراء هذه الاحالات، وفى قاعدة البيانات ما هى قيمة الاحالات وكيفية القيام بها بكفاءة؟ وفى كشافات الكتب هل يستخدم :

التكشيف انظر أيضا التكشيف الآلى: التكشيف المتسلسل

أو التكشيف

الآلى انظر التكشيف الآلى

المسلسل انظر التكشيف المتسلسل

وبالمثل إحالة انظر أيضا

التكشيف ٣٢، ٥٦ انظر أيضا التكشيف الآلى

البحث١٨٣- ١٨٥

التدريب ٧٤

أو

التكشيف ٣٢، ٥٦

البحث١٨٣- ١٨٥

التدريب ٧٤

انظر أيضا التكشيف الآلى

ب. الترتيب: arrangement

هل يكون الترتيب كلمة بكلمة أو حرف بحرف، ومع هذين الاختيارين ما هو نوع المدخل (بالمكان أو بالموضوع) أو يكون الترتيب على أساس الحروف فقط (ليس هناك ما يسمى بالدليل الامبيريقى) وهناك حالات عديدة تتعلق بالمدخل (المكون من عدة كلمات بعضها مجهول أو مبين (مثل كلمة بين Between/ among، تطبيق، من/ فى in /from) فهل يتم تجاهل الحروف والاكتفاء بالاسماء؟.

ج. الطول Length

إلى أى مدى يكون عدد المصطلحات الكشفية أو طول المستخلصات متناسباً مع النص الأصلى؟.

٥- تقييم التكشيف والاستخلاص:

أ. تقييم الاسترجاع:

الطرق المتاحة لتقييم نوعية الاستخلاص تتركز فى منهجية الاستدعاء والدقة Recall and precision وتفريعاتها، وعلى الرغم من مزايا هذه المنهجية إلا أن لها عيوبها الواضحة ولعلنا نذكر منها اثنان هما:

❖ **قياس الاستدعاء:** فى المجموعات الحقيقية يتطلب تقييم كل وثيقة فى المجموعة ومدى صلاحيتها relevance لكل سؤال، وهى مهمة مستحيلة... والدراسات التى تدور حول مشكلة تتصل بالبحث من عدة وجهات نظر، أو البحث بطريقة عريضة Broad ثم تجميع كل الاسترجاعات ذات الصلة relevant لتزويدنا بقاعدة الوثائق الصالحة... هذه الدراسات ليس لديها أى وسيلة لتخمين الوثائق التى يمكن أن تكون ناقصة.

❖ **قياس الصلاحية** relevance measure هو قياس ناقص فى ذاته، فمن الناحية العملية يعامل المقياس كشىء مطلق، ويعتمد هنا على درجة المضاهاة بين ماذا تتحدث عنه الوثيقة Aboutness ومصطلحات السؤال. ولكن هناك عوامل تدخل وتؤثر على فائدة وثيقة معينة فى موقف معين. فعلى سبيل المثال فلا ينبغى ألا تضاهى الوثيقة طبقاً لمستوى معرفة المستفيد، ولكن قيمتها تعتمد جزئيا على ما يعرفه المستفيد مسبقا، فإذا كان هناك وثيقتان متشابهتان، وقد تكون الأولى بالنسبة للمستفيد مفيدة جدا، وتكون الثانية ذات أهمية متدنية لأنها لا تضيف إلا القليل لمعرفة المستفيد أى أن هناك حاجة بحثية ماسة إلى مقاييس أخرى فى المواقع العملية لتحقيق نتائج موثوق بها.

ب. اختبار مدى الاستخدام Usability Testing:

اختبار الاستخدام عملية مكلفة للغاية، وهامش الربح عن مبيعات الكشافات ليس كافيا لدعم هذا الاختبار، ومع ذلك فالبحث فى هذا المجال مفيد للغاية.

ج. قيمة التكشيف:

القضية التى تتعلق بها مختلف الأطراف هى الرد على سؤال هام هو: هل يستحق التكشيف كل هذا الجهد؟ وكيف يمكن قياس القيمة فعلا؟، وإذا كان

التكشيف الإنسانى مكلف فالضغوط مستمرة لتقليل هذه التكاليف.. ولكن ما هى القيمة الفعلية للكشاف، هل كشاف الكتاب يزيد من قيمته؟ وإذا كانت الزيادة صحيحة فكيف يمكن قياسها، والسؤال نفسه ينطبق على النص الكامل والمستخلصات، والأهم هنا أيضا هو حساب القيمة المضافة added value فى حالة وجود الفهرس.. وإذا كان التقييم يتصل بجوهر مشكلة الاسترجاع فهى قضية فى غاية الصعوبة فى الدراسات البحثية.

ثانيا: انتاج وتوليد كشافات الوثائق آليا على مشارف القرن الحادي والعشرين:

أ- تقديم

لقد أدت الزيادة الهائلة فى كمية المعلومات المتاحة من خلال الوسائط الإلكترونية (بما فيها الإنترنت)، إلى زيادة الاهتمام بإنشاء وتطوير طرق آلية للقيام بعملية الاقتباس Extracting أوالترشيح filtering أو التلخيص summarization أو استرجاع المعلومات ذات العلاقة بمهام محددة، ومعظم هذه المعلومات موجودة فى شكل وثائق الكترونية، كالمقالات والتقارير انظر: (Belkin; Cowie; Cardie; Salton) ويلاحظ القارئ استخدام مصطلحات الاقتباس والترشيح والتلخيص فضلا عن الأساليب الفنية التى تمت بالنسبة لاكتشاف المعرفة knowledge discovery فى قواعد البيانات وهى التى تستخدم أيضا عملية اقتباس المعلومات من البيانات النصية، أى أن كلمة مستخلصات واستخلاص لم نعد نستخدم إلا نادرا مع النظم الآلية.

وتشير عملية اقتباس المعلومات إلى اقتباس ملخص المعلومات من النص بحيث تصفه فى شكل تركيبى Structured Form (Cardie, c 1997, P. 65) أما عملية ترشيح المعلومات واسترجاعها فتهتم بتحديد البيانات غير التركيبية أو نصف التركيبية، والتركيز فى عملية ترشيح المعلومات هو على اهتمامات المستفيد user profiles، أما اكتشاف المعرفة فهو مصطلح حديث نسبياً يستخدم مع قواعد البيانات التركيبية (وعادة العلاقية Relational) ويمكن لأساليب اكتشاف المعرفة أن تستخدم مع ا لبيانات النصية أيضا فضلا عن استخدام أساليب اكتشاف المعرفة لتصنيف النص إلى موضوعات Topics (حسب تصنيف مكتبة الكونجرس مثلا...).

ب. أنواع نظم التكشيف الآلية الحديثة وكيفية تقييمها:

قامت الباحثة جوليا هودج (Hodges, J, 2000) باستعراض لأهم مداخل نظم التكشيف الآلية من النواحى الفكرية، ثم اوردت نماذج منها وكيفية تقييمها، ويلاحظ هنا أن تقييمها اعتمد على طرق الاستدعاء والدقة (Recall& precision) على الرغم مما سبقت الإشارة إليه فى خريطة التكشيف إلى سلبيات هذه الطرق التقييمية، وهى تذكر أن هناك نظم تكشيف آلية تستخدم النماذج الاحتمالية Probabilistic models وهى التى تقوم بتقدير الوزن الاحتمالى لمختلف الكشافات، باستخدام عدد من الأسئلة المتعلقة بوثيقة محددة كإحدى الطرق اللازمة للتعرف على الصلاحية Relevance.

وهناك نظم أخرى تجمع بين التوزيع الاحصائى والتحليل اللغوى Linguistic heuristics لتحديد الكشافات الصالحة للوثيقة، وهناك مداخل أخرى تفيد من تجهيزات اللغة الطبيعية (N L P) بالتعرف على جذور الكلمات Stemming (أحيانا نسمى الجذر Root والجذع Stem) كما تحاول نظم أخرى الإفادة من معلومات المتن (أو المضمون) Contextual information لتكشيف الوثائق.

وترى جوليا هودج Hodges أن معظم التركيز الحالى فى بحوث التكشيف الآلى هو تكشيف مفاهيم Conceptual Indexing (أى ما يسمى أيضا بالتكشيف بالتعيين assignment indexation) فى مقابل نظم التكشيف المعتمدة على الكلمات word indexing من النص عن طريق الاقتباس Extraction... وهى تؤكد على ضرورة أن تكون نظم التكشيف ممثلة لدلالات semantics محتوى المقال، وليس مجرد الطرق التركيبية syntactical ways، ذلك لأن التكشيف المفهومى يساعد عملية استرجاع المعلومات لسببين على الأقل اولهما أن هناك مفاهيم أقل من امكانيات التمثيل التركيبى syntactic representation وثانيهما أن الكشافات المفهومية يمكن أن تقدم لنا المعلومات الدلالية semantic الهامة، وذلك عندما يحاول نظام استرجاع المعلومات مضاهاة السؤال بالمعلومات فى الوثيقة. ولكن التكشيف المفهومى يحتاج إلى المعرفة الواسعة بالتخصص العلمى فى المجال. وقد قامت جوليا هودج باستعراض نماذج من نظم التكشيف المختلفة ومداخلها البحثية وهى

١- نظام سمارت SMART (Salton Magic automatic retriever of text).

٢- نظام ايمز AIMS (assisted indexing at Mississippi state).

٣- نظام ديدال Dedal وهو كشاف مفهومى للنص والرسومات ووثائق التسجيلات videotaped فى حقل الهندسة الميكانيكية.

٤- نظام وورلد فيو World View وهو يقوم بتجهيز ومعالجة مقالات الأخبار والتقارير الفنية من المؤسسات المختلفة مثل معامل بيل Bell lab.

٥- نظام سيزور Scisor System

(System for Conceptual information summarization Organization and retrieval) وهو يقوم بتحليل القصص الإخبارية خصوصا الأخبار المالية.

٦- نظام كوسكوت Quescot System

(Quantification and encapsulation of semantic context) واعتمدت فى التقييم على مقاييس الاستدعاء والدقة حيث:

الاستدعاء = عدد الكشافات الصحيحة التي تم توليدها / العدد الكلى للكشافات الصحيحة

الدقة = عدد الكشافات الصحيحة التي تم توليدها / العدد الكلى للكشافات التى تم توليدها

وقد كان أفضل معدل استدعاء ودقة تم الوصول إليها عن طريق خبراء ما يسمى بمؤتمر التفاهم للرسالات (MUC) هو حوالى ٤٠%- ٥٠% (Cowie, J, 1996)، وهذه نظم تقوم بالاقتباس والتلخيص.. ولكن نظام ايمز AIMS كان له معدل دقة تصل إلى (٦٢- ٦٦,٥%) ومعدل استدعاء (٧٢- ٩٠%) (Hodgs, J., 2000, P. 15).

• وقد استخدم النظام عدة تجارب تفيد من مختلف أجزاء المقال • المستخلص وحده • المستخلص والمقدمة • المستخلص وجزء من المقدمة • المستخلص والخاتمة Conclusion • المستخلص وجزء من المقدمة وجزء من الخاتمة.

ثالثا: تحليل لبعض التطورات المعاصرة للتكشيف والاستخلاص فى الإنتاج الفكرى:

اختار الباحثان عشرين مقالا من مجلة الاتجاهات المكتبية (١٩٩٩) Library Trends ومن موسوعة المكتبات والمعلومات ELIS عام (٢٠٠٠م) ومن مجلة الجمعية الأمريكية لعلم المعلومات JASIS ومعروف أن الدورية الأخيرة هى أهم الدوريات فى مجالها على المستوى العالمى. ولعل استعراض المقالات السابقة والمنشورة ما بين الأعوام ١٩٩٧/ ٢٠٠٠م يظهر لنا أضواء وجوانب بعض التطورات المعاصرة والمستقبلية فى التكشيف والاستخلاص:

أ. نماذج من البحوث والدراسات بمجلة الجمعية الأمريكية لعلم المعلومات (JASIS): وتشمل هذه النماذج عنوان البحث ونبذة عنه والوحدة العلمية القائمة بالبحث :

1- Disciplinary Variation in Automatic Sublanguage Term Identification

١- يصف البحث طريقة لتحديد مصطلحات المجال الموضوعى الأساسية والموجودة باللغة البديلة (SL) وبالطريقة الآلية، وتعرف اللغة البديلة (SL) بأنها اللغة المستعملة فى مجال موضوعى محدد بواسطة المتخصصين.

- مدرسة علم المعلومات والمكتبات جامعة شمال كارولينا (شابيل).

2- A Statistical Learning Approach to Automatic Indexing of Controlled Index Terms.

٢- مدخل إحصائى تعليمى للتكشيف الآلى الخاص بالمصطلحات المحكومة للكشافات ويهدف البحث إلى تحسين أداء عملية التعلم، وهذه بدورها ستؤدى إلى تحسين أداء التكشيف.

- قسم علم الحاسب الآلى الجامعة الصينية فى هونج كونج

3- A Concept Space Approach to Addressing the Vocabulary Problem in Scientific Information Retrieval: An Experiment on the

Worm Community System.

٣- استخدام مدخل سعة المفهوم لحل مشكلة المصطلحات فى استرجاع المعلومات العلمية: تجربة على نظام مجتمع الديدان مع التعرف على استخدامات هذا المدخل فى شبكة الإنترنت والمكتبات الرقمية.

- قسم نظم المعلومات الإدارية (MIS) مدرسة الإدارة جامعة اريزونا
- المدرسة العليا لعلم المكتبات والمعلومات جامعة الينوى (اريزونا- شامبين)

4- Why words and co- words Cannot Map the Development of the Sciences.

٤- لماذا لا تؤدى الكلمات والكلمات المصاحبة إلى وضع خريطة لتطور العلوم

أثبت الباحث رسالته هذه باستخدام مقالات النص الكامل فى مجال الكيمياء الحيوية وذلك لتحليل الكلمات ومقارنتها بالنسبة لتواجدها أو غيابها المصاحب فى هذه النصوص.

- قسم ديناميكية العلوم والتقنية جامعة امستردام هولندا.

5- Citation Context Versus the Frequency Counts of Citation Histories.

٥- سياق الاستشهاد مقابل حساب التكرار لتواريخ الاستشهاد، وقد استخدم الباحثون أكثر من مائتى ورقة بحث لتحليلها.. وتبين للباحثين أنه ليس هناك تطابق بين نتائج رتبة سياق الاستشهاد عند مقارنتها برتبة موضوعات البحوث عن طريق حساب تكرار استشهاداتها.

- وحدة دراسات العلوم - كلية الطب جامعة زغرب، كرواتيا.

6- Professional Summarizing: No Cognitive Simulation Without Observation.

٦- التلخيص المهنى: ليس هناك محاكاة معرفية بدون الملاحظة.. يشير الباحث إلى أن التلخيص المهنى يتضمن عمليات معرفية كالاستخلاص والتكشيف والتصنيف والتى يقوم بها الخبير لاعداد نظم المعلومات الببليوجرافية والتى تمثل صيغاً مصغرة للقارئ، وتعتبر مرشحات للمعلومات، ولكن العمليات المعرفية تتم داخل رؤوسنا وبالتالى فهى تستبعد الملاحظة المباشرة، وينادى الباحث بضرورة دراسة التجهيزات الفنية والمنهجية اللازمة للإفادة من الملاحظة.

- قسم المعلومات والاتصال معهد البوليتكنيك- هانوفر- المانيا

7- Khan, Kushal and Craig Locatis (1998, 172- 182) Searching through Cyberspace: The Effects of Link Display and Link Density on Information Retrieval from Hypertext on the World Wide Web (JASIS, 49 (2)).

٧- البحث خلال الفراغ الكونى: تأثير روابط العرض والكثافة على استرجاع المعلومات من النص الفائق (الهيبرتكست) على الشبكة العنكبوتية العالمية:

تتناول الدراسة استرجاع المعلومات من الهيبرتكست على الشبكة العنكبوتية العالمية للإنترنت، وقد تبين وجود تأثيرات ذات دلالة لكل من كثافة الروابط (عدد الروابط لكل عرض) وشكل العرض (فى الفقرات والقوائم) وذلك بالنسبة لنتائج الاداء فى البحث، كما يعتمد الاداء على الخبرة المسبقة فى تصفح واستعراض الهيبرتكست على الشبكة العنكبوتية.

- قسم العلم المعرفى (Cognitive Science) المكتبة الوطنية الطبية- أمريكا.

8- Khan, Kushal (1998) Searching though Cyberspace: the effects of Link Cues and Correspondence on information Retrieval from Hypertext on the world wide web. (JASIS, 49) (14) 12 48-1253).

٨- البحث خلال الفراغ الكونى: تأثير الروابط المعبرة وروابط المراسلة على استرجاع المعلومات من النص الفائق (الهيرتكست) على الشبكة العنكبوتية العالمية.

تتناول الدراسة تأثير الروابط المعبرة وروابط مراسلة الكلمات على الاداء البحثى باستخدام وثيقة هيرتكست محورية ولها روابط خارجية بمصادر أخرى على الشبكة العنكبوتية العالمية.

- جامعة الينوى والمكتبة الوطنية الطبية.

9- Clumping Properties of Content- Bearing Words

٩- الصفات التجميعية للكلمات الحاملة للمحتوى الموضوعى المميز

تعتمد هذه الدراسة على ما تقوم به نظم استرجاع المعلومات من التعرف على الكلمات التى تحمل محتوى موضوعى مميز، كما قد تعين هذه النظم أوزاناً لهذه الكلمات كجزء من عملية صياغة الأسئلة ... وتحدد هذه الدراسة فكرة التجمع المتسلسل Clustering للكلمات فى النص كمؤشر للمحتوى الذى تحمله الكلمات... وقد تم تجميع الادلة التجريبية من قواعد معلومات النص الطبيعى ومن ثلاث لغات مختلفة.

- مركز دراسات المعلومات واللغات — جامعة شيكاغو- أمريكا
- قسم الرياضيات — جامعة بارلان Bar LLan إسرائيل
- قسم علم الحاسب — جامعة توركو Turku فنلندا

10- Price Index and Its Relation to the Mean and Median Reference Age.

١٠- كشافات برايس وعلاقته بمتوسط ووسيط العمر المرجعى

يتضمن المقال فى جزئه الأول اثبات أن كشاف برايس يعتبر دالة لمتوسط ووسيط العمر المرجعى، أما الجزء الثانى فيشير إلى العلاقة المحددة والتى يتم شرحها بواسطة نظرية الاحتمالات والاحصاء.

- جامعة ليمبرج Limburg — بلجيكا

11- Interaction in Information Retrieval: Selection and Effectiveness of Search Terms.

١١- الروابط البينية فى استرجاع المعلومات: اختيار وفاعلية مصطلحات البحث:

تناولت الدراسة مصادر وفاعلية مصطلحات البحث المستخدمة اثناء البحث على الخط المباشر وبوجود وسيط معلوماتى، وكان عدد المستفيدين (أربعين) وعدد الوسطاء (أربعة) حيث كان المستفيد يقدم سؤالاً ويتفاعل مع الوسطاء الأربعة وذلك باستخدام نظام ديالوج. وكانت أكثر المصادر فاعلية تلك التى تعتمد على مصطلحات البحث المكتوبة فى أسئلة الباحثين فضلاً عن المصطلحات المتعلقة بالتغذية المرتدة.

- قسم علوم المكتبات والمعلومات جامعة شمال تكساس- أمريكا
- مدرسة دراسات الاتصال والمعلومات والمكتبات جامعة راتجرز

12- Design and Implementation of Automatic Indexing for Information Retrieval With Arabic Documents.

١٢- تصميم وتطبيق التكشيف الآلى لأغراض استرجاع المعلومات باستخدام الوثائق العربية

استخدمت فى الدراسة عدد (٢٤٢) مستخلص للوثائق العربية فى مجال نظم المعلومات وعمل الحاسب الآلى، كما تم تصميم وبناء نظام استرجاع معلومات آلى لمعالجة البيانات العربية، واستخدم النظام لغة (C) والمترجم (GCC) على IBM/ PC وحاسب مصغر متلائم، واثبتت التجربة أن التكشيف الآلى له نفس فاعلية التكشيف اليدوى.

- معهد الأردن للعلوم والتقنية اربد- الأردن
- قسم علم الحاسب الآلى معهد الينوى للتقنية- شيكاغو

13- Experiments with Automatic Indexing and a Relational Thesaurus in a Chinese Information Retrieval System.

١٣- تجارب فى التكشيف الآلى والمكنز العلاقى فى نظام استرجاع المعلومات الصينى

تجارب مع نظام استرجاع تفاعلى (CIRS) ومكنز علاقى تفاعلى... تقاس الفاعلية بواسطة الاستدعاء والدقة وقد تبين بالتحليل الاحصائى أن قياسات الاستدعاء والدقة تتحسن بواسطة المكنز العلاقى سواء كان ذلك فى التكشيف الآلى أو اليدوى.

- قسم علم الحاسب الآلى   معهد الينوى للتقنية- شيكاغو

14- Indexing and Access for Digital Libraries and the Internet: Human, Database, and Domain Factors.

١٤- التكشيف والاتاحة للمكتبات الرقمية والإنترنت: العوامل الإنسانية وقواعد المعلومات والمجال العلمى

دراسة لبعض العوامل والقضايا التى تؤثر على التكشيف والاتاحة عند تصميم النظم لاسترجاع المعلومات فى المكتبات الإلكترونية الرقمية المستقبلية والإنترنت. فالمستفيدون يبحثون فى كيفية الوصول إلى أفضل النتائج على اسئلتهم عند البحث على الشبكة العنكبوتية العالمية خصوصاً وهناك من الباحثين القريبين لعلم المعلومات (كالعلم المعرفى واللغويات المحسبة والذكاء الاصطناعى) ممن يدرسون قضايا استرجاع المعلومات من وجهة نظرهم. وتشمل العوامل الإنسانية: البحث الموضوعى بالمقارنة بالتكشيف والمصطلحات المتعددة للاتاحة، وكذلك التصنيف الذى يتبعه الشخص العادى Folk Classification أما عوامل قاعدة البيانات فتشمل قانون برادفورد والمصطلحات المساعدة المرتبة Vocabulary Scalability وقاعدة ريزينكوف Resnikoff- Dolly 30: 1 Rule أما عوامل المجال العلمى فتشمل دور المجال فى التكشيف.

- قسم دراسات المعلومات/ جامعة كاليفورنيا، لوس انجلوس

ب. من مجلة الاتجاهات المكتبية (عام ١٩٩٩) :

15- He, Qin (Summer 1999)

Knowledge discovery through Co- Word Analysis. Library

Trends. V 48 (1), P. 133- 159.

١٥- اكتشاف المعرفة من خلال تحليل الكلمات المصاحبة وتتناول الدراسة تحليل مصاحبة الكلمات كاسلوب للتعريف على عدد مرات الحدوث المصاحب المشترك Co- Occurrence لزوج من الكلمات أو الفقرات، وذلك لاكتشاف الروابط بين موضوعات البحوث وبالتالى تتبع تطور العلم فى مجال معين، وذلك يؤدى إلى التخطيط الأفضل للبحوث العلمية الأمر الذى يجعل هذا الأسلوب اداة قوية لاكتشاف المعرفة في قواعد البيانات.

16- Qin, Jian (Summer 1999) Discovering Semantic Patterns in Bibliographically Coupled Documents. Library Trends, V. 48 (1), p. 109- 132.

١٦- اكتشاف النماذج الدلالية فى الوثائق المصاحبة ببليوجرافيا

تتناول الدراسة قضايا اكتشاف المعرفة فى قواعد البيانات الببليوجرافية، وقد استخدم مثال التحليل النموذج الدلالى للتعريف بالجوانب المنهجية لاكتشاف المعرفة فى قواعد البيانات اعتمادا على الكلمات المفتاحية المختارة من الوثائق والمجمعة بواسطة المصاحبة الببليوجرافية.

17- Ahonen, Helena (Summer 1999) Knowledge Discovery in Documents by Extracting Frequent Word Sequences. Library Trends, V. 48 (1), p. 160- 181.

١٧- اكتشاف المعرفة فى الوثائق عن طريق اقتباس الكلمات ذات التكرار العالى

تتناول الدراسة طريقة لاقتباس أكثر الكلمات المتتابعة ترددا فى الوثائق حيث تحاول طرق اكتشاف المعرفة التعرف على النماذج والانتظامات العامة فى البيانات بدلا من الكشف عن حقائق معينة... وهذه الطريقة عسيرة للتطبيق بالنسبة للإنسان ولكنها ممكنة آليا.

18- Pinto, Marra and Lancaster, F. W. (Summer 1999). Abstracts

and Abstracting in Knowledge Discovery. Library Trends V. 48 (1), P. 234- 248.

١٨- المستخلصات والاستخلاص وأهمية فى اكتشاف المعرفة تتناول هذه الدراسة متطلبات المستخلصات التى يمكن قراءتها بواسطة الإنسان ومقارنتها بالمتطلبات الخاصة بالبحث بالحاسب الآلى، وينتهى الباحثان إلى أن الإتاحة الواسعة للنص الكامل الالكترونى، لا يقلل من قيمته أنشطة المستخلصات لاسترجاع المعلومات حتى فى التطبيقات الأكثر تعقيدا فيما يعرف الآن باكتشاف المعرفة .Knowledge Discovery

19- Chowdhury, Gobida G. (Summer 1999) Template Miningy for information Extraction from Digital Documents, Library Trends, V. 48 (1), P. 182- 208.

١٩- استخراج اقتباسات المعلومات من الوثائق الرقمية.

تتناول الدراسة عملية اقتباس المعلومات آليا من نصوص اللغة الطبيعية اعتمادا على مدخل الوسادة Template وذلك عندما تشكل البيانات او النص المحيط بها نماذج يمكن التعرف عليها، ويستخدم فى هذه العملية محركات بحث الوب Search Engines مثل التافيزتا وذلك لمعاونة المستفيد النهائى فى توليد الاقتباسات من الوثائق الرقمية.

<u>ج . نموذج من بحوث عام (٢٠٠٠م) فى موسوعة علم المكتبات والمعلومات</u>

20- Hodges, Julia (2000) مقالة الموسوعة

Automated System for the Generation of Document Indexes. In: Encyclopedia of library of information Science V. 66, Suppl. 29, PP 1- 19.

٢٠- نظام آلى لتوليد كشافات الوثائق

يتناول هذا المقال الاستعراضى المطول إنتاج الكشافات بنظام آلى، ويركز على أن البحوث المعاصرة فى التكشيف الآلى هى بحوث فى التكشيف المفهومى

Conceptual Indexing بالمقابلة بنظم التكشيف التى تعتمد على الطرق (التركيبية أو الصرفية) التى تركز الاعتماد على الكلمات فى النص، أى أن التركيز الحالى هو على الكشافات التى تمثل دلالات Semantics المحتوى فى المقال، وليس على الطرق التى يتم بها تمثيل تلك المفاهيم نحويا فى النص... ويركز مقال الموسوعة أعلاه على كيفية إنتاج الكشافات الآلية للوثائق فى شكل مقروء آليا وباللغة الطبيعية.

رابعا: اتجاهات فى النظرية والممارسة :

١- من الإنتاج الفكرى المطبوع إلى الإنترنت:

تناول الباحث وايلى (Wilwy, D. L. 1994) التطورات التى حدثت بالنسبة لخدمات التكشيف والإستخلاص وتساءل هل هناك مستقبل لهذه الخدمات التقليدية المطبوعة؟ وفى إجابته عن هذا التساؤل ذهب وايلى إلى أن العديد من نظم المعلومات تواجه اليوم خطر التقادم نظراً للتطورات السريعة فى تكنولوجيا الحاسبات والاتصالات، وأن بعض المنتجين لقواعد البيانات العلمية الرئيسة يقومون فى الوقت الحاضر بتطوير نظم إنتاجية جديدة للاستجابة للاحتياجات المتغيرة للسوق، ولمتطلبات الطرق السريعة للمعلومات..

وإذا كان الماضى القريب قد شهد اندماج مختلف التكنولوجيات فى نهاية الستينات وأوائل السبعينات بما فى ذلك نظم اختزان البيانات المنخفضة التكاليف ونظم الاتصالات الرخيصة، فضلاً عن تحسيب العديد من عمليات النشر... التى أدت بدورها لتطوير نظم البحث على الخط المباشر.. فإن هناك اندمجاً مماثلا للجيل التالى من التكنولوجيات المتمثل فى الحاسب الشخصى المنخفض التكاليف ذى القوة الاختزانية الكبيرة فضلا عن نظم الاتصالات السريعة للغاية المسماة بالزبون/ الخادم Client/ Server وكذلك النشر الإلكترونى على المكتب، وهذه أدت بنا إلى الطرق السريعة للمعلومات..

كما يشهد الوقت الحاضر العديد من الهيئات التى استكملت إنشاء "النظم الإنتاجية الجديدة" وحولت الهيئات محور اهتمامها من نظام إنتاج للمطبوعات إلى إنتاج مواز للأشرطة اللازمة للبحث على الخط المباشر ثم إلى إنتاج ملفات الخط المباشر والتى يمكن أن تنتج بدورها أشكالا من المطبوعات والمنتجات الإلكترونية... وهذا مفهوم

بسيط، ولكنه ضخم التعقيد عند التنفيذ.. وقد لوحظ أن الشكل الإلكترونى للبيانات قد أدى إلى دخل أكثر ويستمر فى الزيادة.

ولكن ماذا يحمل المستقبل؟ الناشرون يرون طلبات أكبر على بياناتهم فى الشكل الإلكترونى وأنهم فى حاجة إلى مرونة لتوصيل المنتجات التى يحتاجها المستهلك أمام المنافسة المتزايدة فى إنتاج المعلومات الإلكترونية. ولذلك كان الاتجاه نحو الإنترنت (Wiley, D. L, 1994).

٢- النظم الإنتاجية الجديدة لخدمات التكشيف والاستخلاص:

تنشئ خدمات الاستخلاص والتكشيف نظماً إنتاجية جديدة طبقاً لخطوات منهجية عديدة تبدأ بتشكيل لجان استشارية للتعرف على مختلف جوانب النظام المطلوب من التحرير إلى التسويق إلى التحسيب إلى الإدارة، وعلى سبيل المثال لا الحصر، فقد قامت هيئة المستخلصات الكيميائية الدوليةCABI بتعميم المعلومات عن نظامها الإنتاجى الجديد، وكان عليها أن تتعامل مع تاريخ طويل لعدد (٢٦) دورية مستخلصات مستقلة، وقد كان لديها قاعدتين للبيانات يتم إنتاجهما فى مكانين مختلفين وتستخدمان ممارسات تكشيف مختلفة ثم قامت CABI بتحويلها إلى مدخلات لقاعدة بيانات مركزية للمستخلصات الكيميائية، والتى يمكن استخلاص مختلف المنتجات منها.

لقد شكلت هيئة CABI لجنة ممثلة لمختلف الأنشطة المتصلة بالتحرير وضبط الجودة والنظم والتسويق والتزويد من أجل تصميم النظام الجديد... ومع ذلك ومع وجود أكثر من مائة اختصاصى معلومات يقومون بالعمل فى انتاج قاعدة البيانات فقد كان ذلك تحدياً فى تصميم النظام الذى يرضى الجميع.. وتبنى الجميع مفهوم "التكشيف من أجل قاعدة البيانات" وليس التكشيف لدورية متخصصة بعينها.. وقد أدى ذلك إلى تكشيف أكثر انتظاماً بالنسبة للباحثين للقاعدة على الخط المباشر، وأدى ذلك التنظيم إلى سهولة إنتاج منتجات جديدة للمستهلكين.

لقد قامت هيئة المستخلصات الكيميائية بإعادة تكشيف قاعدة بيانات المستخلصات الكيميائية والتى تحتوى على أكثر من ٢,٨ مليون تسجيلة.. وقد تم التغيير الفعلى بواسطة برامج الحاسب الآلى، ولكن سبقها جهد ضخم يتصل بتقرير وتحديد ماذا يتم تغييره وكيفية هذا التغيير؟

فبالإضافة إلى تصحيح البيانات القديمة، أضيفت حقول جديدة إلى البيانات الداخلة فى النظام، وذلك لضمان جودة نوعية المستخلصات، وتوفير نقاط إتاحة إضافية للمستفيد. ومن الحقول الجديدة التى أضيفت هناك العنوان البريدى للمؤلف Author's e- mail إذا كان فى المقال الأصلى وكذلك بعض مصطلحات التكشيف الإضافية للاسترجاع الأكثر دقة.. كما قامت CABI بإضافة حقول جديدة شاملة لحقل الواصفات حتى تمكن المستفيدين من البحث بالمصطلحات العريضة فى مكنز المستخلصات الكيميائية بطريقة هرمية، ثم الاسترجاع الآلى للمصطلحات الأضيق دون الحاجة إلى اختيارها بطريقة فردية، كما أضافت فوق هذا كله أكواد تصنيفية جديدة لكل تسجيلة فى قاعدة البيانات.

٣- التغييرات المعاصرة والمستقبلية والمستفيد:

لقد كانت عين المسؤولين على المستيفيد وكيفية توصيل هذه المتغيرات إليه.، خصوصاً وأنه فى عصر فيضان المعلومات الغامرة INFORMATION OVERLOAD، وبالتالى فإن معظم المستفيدين لا يعرفون كيفية بناء قاعدة البيانات ولا كيفية استرجاع المعلومات بالطريقة السليمة فهم يطبعون الكلمات على النهاية الطرفية ثم يتقبلون أى مخرجات يحصلون عليها. ومع ذلك فينبغى التأكيد على أننا فى عصر المعلومات نستطيع أن نبنى أدوات استرجاع بالغة التعقيد وجعلها سهلة الاستخدام بالنسبة للمستفيد، وعلى كل حال فيجب أن يكون المستفيد قادراً على طباعة كلمة او مصطلح فى تخصصه ثم يقوم النظام بطريقة آلية بالبحث عن المصطلحات القريبة أو الشبيهة، تغييرات تتصل بكيفية الهجاء بكيفية الهجاء أو كونها جمع أم مفرد.. الخ دون اسهام منه فى ذلك.

ونظم الاسترجاع هذه يمكن أن تتطور فقط مع التكشيف الدقيق والتيجان التى تتوفر فى قاعدة البيانات كإحدى مكوناتها.. وما هو سهل بالنسبة للمستفيد فى الاستخدام يعتبر عادة معقد للغاية بالنسبة لأولئك الذين قاموا بالإعداد وإنتاج قاعدة البيانات.

٤- إدخال البيانات مفتاح إنتاج قاعدة البيانات الجيدة:

تقوم جميع النظم الجديدة بإدخال بعض الأساليب المستخدمة لإدخال البيانات، ومعظم هذه النظم تحاول اكتشاف تكنولوجيات إضافية للمستقبل.. وقد شرح لنا

أولارى (O' LEARY, M., 1990) ذلك عندما قال، إن إدخال البيانات بواسطة لوحة المفاتيح مازال الطريقة السائدة، ومع ذلك فهناك نظم عديدة قد تحولت إلى الإدخال المباشر بواسطة اختصاصى المعلومات الذى يقوم بالتكشيف والاستخلاص. ولقد كانت المرحلة الأولى لتطبيق نظام الإنتاج الجديد للمستخلصات الكيميائية CABI هو وضع حاسب شخصى على مكتب كل واحد من المستخلصين.. وقد وفر ذلك عملية مرهقة خاصة بكتابة المستخلص باليد ثم إدخالها بلوحة المفاتيح ثم تحريرها ومراجعتها وتصحيحها.

أما فى نظام المعلومات الوطنى NTIS فقد طور نظاما معقداً للغاية حتى يكون قادراً على تناول مختلف الأشكال FORMATS وتزويد مختلف المصالح الحكومية بالمعلومات الإلكترونية، ويركز النظام الوطنى NTIS على شبكة الإنترنت وغيرها من الوصلات لمصادر المعلومات الحكومية.

أما فى معهد المعلومات العلمية ISI فهم يرون التصوير IMAGING كموجة المستقبل وذلك حتى تقدم خدمة إحاطة جارية وتزويد الوثائق والاختزان والتوصيل الإلكترونى للوثائق باعتبارها نشاطاً محورياً للنجاح فى المستقبل، وإلى جانب تركيز معهد المعلومات العلمية على تكنولوجيا التصوير فهو يهتم بالاتصال بالناشرين الذين يقومون بالنشر الأولى لأخذ الحقوق الإلكترونية منهم ووضع نظام قانونى LICENSING SYSTEM بالاتفاق مع الناشرين ويقول O' LEARY أنه حتى عام ١٩٩٠ كان هناك حوالى ٥٧% من الــ (١٩٠٠) ناشر الذين تم توقيع اتفاقات معهم للحقوق الإلكترونية، ومن الملاحظ أن معهد المعلومات العلمية لديه (٢٧٨٤) ناشر للدوريات التى تحتويها منتجاته، كما قام المعهد بعقد توصيل الوثائق مع حوالى ٧٠% منهم.

ولكن المستقبل بعد هذا كله مازال غامضاً، فماذا سيكون عليه نظام توصيل المعلومات المستقبلى؟ وما هو دور الناشرين للإنتاج الفكرى الأولى وللإنتاج الفكرى الثانوى (التكشيف والاستخلاص) وخدمات توصيل الوثائق، خصوصاً والمؤلف سيمكنه النشر مباشرة على الإنترنت، كما سيقوم العاملون بتوصيل الوثائق بإنشاء قواعد معلومات مجانية لتسويق خدماتهم، وسينتج الناشرون قائمة محتويات إلكترونية للإنتاج الفكرى الأولى (الدوريات والكتب)، كما سيبحث الناشرون للإنتاج

الفكرى الثانوى، إمكانية إعداد النص الكامل للوثائق فضلاً عن التكشيف والاستخلاص، والوقت الحاضر هو وقت إعادة النظر فى تقييم خدمات التكشيف والاستخلاص الحالية.. ذلك لأنه فى هذا العصر الذى يتميز بتوفر النص الكامل يطرح السؤال التالى:؛ هل هناك حاجة إلى خدمات التكشيف والاستخلاص؟.. ويذهب الباحث أولارى O' Leary إلى أن هذه الخدمات يمكن أن تستمر ما دامت تستقوم بالتحديث ومراجعة خدماتها بحيث تقدم للمستهلك ما يريد، أى تقديم نوعية عالية من المنتجات المكشفة على مستوى عال أيضا، وتقديم البيانات فى مختلف الأشكال.. فالهيئات التى ستعيش فى المستقبل هى تلك التى ستعرف حاجة السوق جيداً وتستطيع الاستجابة السريعة لتجميع المعلومات وتنظيمها وترشيحها للمستفيدين.

٥- توقعات لانكستر وتنبؤاته:

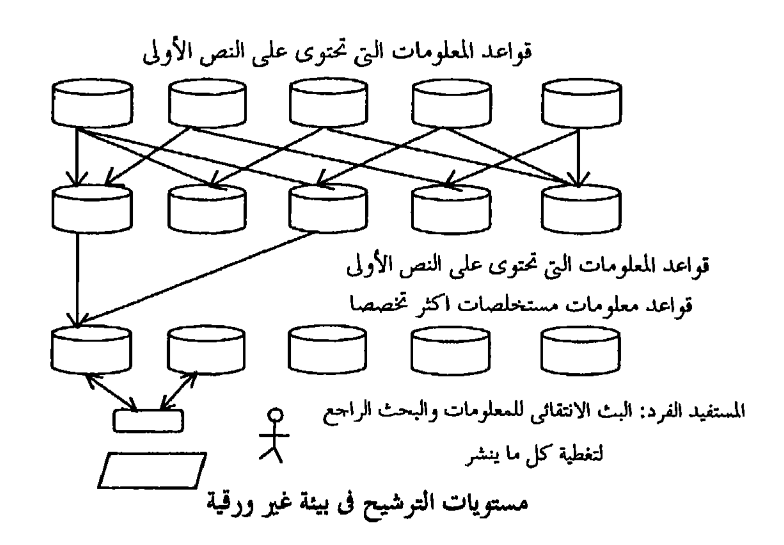
ولعلنا نختم هذه التوقعات ببعض التنبؤات التى أوردها لانكستر فى نهاية كتابه عن التكشيف والاستخلاص (Lancaster, F., 1991) وكتابه الأحدث عام (١٩٩٨) بنفس العنوان.

الهجرة من الورق إلى الإلكترونيات وبعض التوقعات المستقبلية:

قام الباحثان لانكستر وجولدهور بإرسال استبيان إلى عدد (٢٠٠) مكتبة فى المصالح الحكومية والصناعة والجامعات وذلك لتقدير مدى هذه الهجرة إلى الإلكترونيات.. وقد اكتشفا قيام العديد من هذه الهيئات بإلغاء اشتراكاتهم فى خدمات التكشيف والاستخلاص، وإن كان ذلك يعزى أيضاً لانخفاض ميزانيات المكتبات وانخفاض القوة الشرائية.. كما اكتشفا أن المكتبات الحديثة الإنشاء تتجه مباشرة إلى الإتاحة على الخط المباشر دون اللجوء إلى المرحلة الإنتقالية وهى المصادر المطبوعة.

هذا ويحتوى الإنتاج الفكرى على العديد من التوقعات أو التنبؤات الخاصة بمستقبل خدمات التكشيف والاستخلاص ومن بينها توحيد النشر الأولى والثانوى خصوصاً والظروف المعاصرة تشجع على ذلك .. حيث يمكن أن تكون الخدمات الثانوية الببليوجرافية (خدمات الإتاحة) مستمدة آلياً من المواد الأولية الإلكترونية، وأن تكون متاحة فى نفس الوقت الذى يصدر فيه الإنتاج الفكرى الأولى ... وبالتالى فيمكن للمستفيدين الانتقال من أدوات الإتاحة هذه إلى النص الكامل للوثائق

على نفس النهاية الطرفية على الخط المباشر. بل ويتوقع لانكستر الاختفاء الكامل لمعظم ـ إن لم يكن كل ـ الخدمات الثانوية، ومع ذلك فهو يتوقع زيادة الاهتمام بالمستخلصات الإلكترونية (Kuhlen, 1984) كآلية هامة لترشيح الاسترجاع المحسب ونظم البث الانتقائي للمعلومات ولعل الشكل التالي أن يعكس هذا التطور. (Lancaster, F., 1991, p. 260)



مستويات الترشيح في بيئة غير ورقية

خامسا: بعض النتائج والتوصيات:

١- على الرغم من التناقص التدريجي لبحوث التكشيف والاستخلاص إلا أن البحوث النظرية المتصلة بنظم التكشيف والاستخلاص وتطبيقاتها قد زادت تعقيداً خصوصا بالنسبة للنظم المفهومية Conceptual System فضلا عما ورد في الدراسة عن خريطة البحوث ومجالاتها الخمسة (انظر الجزء الأول والثاني من الدراسة)

٢- **التعقد الموضوعي في مجالات البحوث:** لعل المصطلحات الجديدة في العناوين وفي المتن والتي لا يتوفر لها ترجمة مقبولة باللغة العربية تعكس هذا التعقد، فضلا عن استخدام الاحصاء والرياضيات والحاسبات في معالجة هذه الموضوعات المعقدة (انظر الجزء الثالث من الدراسة).

٣- فى المنهج والأدوات: يستخدم المنهج التجريبى فى العديد من هذه البحوث، ومعروف ندرة استخدام هذا المنهج فى دراسات المكتبات والمعلومات بصفة عامة.

٤- فى بحوث الفريق: معظم هذه البحوث قام بها أكثر من باحث واحد من تخصصات وأقسام علمية مختلفة وعلى سبيل المثال لا الحصر (قسم الحاسب الآلى/ قسم ديناميكية العلوم والتقنية/ قسم نظم المعلومات الإدارية/ قسم العلم المعرفى/ قسم الرياضيات/.... إلى جانب أقسام المعلومات والمكتبات والاتصال).

٥- بحوث بين الجامعات داخل الدولة وعبر القارات: (الملفت للنظر أن بحوث الفريق هذه تتم فى حالات عديدة بين جامعات مختلفة داخل القطر الواحد وبين جامعات فى أقطار مختلفة وعلى سبيل المثال لا الحصر (جامعة اريزونا جامعة الينوى بأمريكا)، (جامعة شيكاغو وجامعة بارلان بإسرائيل وجامعة توركو بفنلندا)، (معهد الأردن للعلوم والتقنية ومعهد الينوى للتقنية بشيكاغو..) ... الخ

ولعل أهم التوصيات التى يمكن أن يوردها الباحثان، هو ضرورة زيادة وعمق المقررات فى التكشيف والاستخلاص التى تقدمها أقسام المكتبات والمعلومات العربية، وأن تنشأ وحدة بحوث التكشيف والاستخلاص ضمن مركز نظم المعلومات بكلية آداب جامعة القاهرة يستقطب فيها بعض علماء اللغويات والمعلومات والحاسبات للمعاونة فى استيعاب هذه التطورات الكبيرة فى المجال وإمكانية تطويع اللغة العربية لاستخدامها.

مراجع عربية

١- أحمد بدر: (١٩٦٤) التوثيق الآلى، ثورة فى عالم المكتبات. **مجلة المكتبة العربية**، القاهرة، مج١، ع٤، ص ١٨-٢٦.

٢- أحمد بدر: (١٩٩٢). مصادر المعلومات فى العلوم والتكنولوجيا. ـ الرياض: دار المريخ، ٣٩١ص .

٣- أحمد بدر: (١٩٩٦) أساسيات فى علم المعلومات والمكتبات. الرياض: دار المريخ.

٤- أحمد بدر: (١٩٩٨) تقييم الكشافات والتكشيف. **مجلة المكتبات والمعلومات العربية**، الرياض، س١٨، ٤ ١، ص ٤٧-٦٧.

٥- حشمت قاسم: (٢٠٠٠) **مدخل لدراسة التكشيف والاستخلاص**. القاهرة: دار غريب.

٦- حشمت قاسم: (١٩٨٤). خدمات المعلومات: مقوماتها وأشكالها. ـ القاهرة: مكتبة غريب للطباعة والنشر، ص ص ٢٢٢-٢٣٤.

٧- كنت، ألن: ثورة المعلومات، استخدام الحاسبات الإلكترونية فى اختزان المعلومات واسترجاعها. ترجمة حشمت قاسم وشوقى سالم ومراجعة أحمد بدر، وكالة المطبوعات، الكويت، ١٩٧٣.

٨- لانكستر، فردرك: ولفرد. (١٩٩٧) أساسيات استرجاع المعلومات/ فردرك ولفرد لانكستر، آمي. وورنر؛ ترجمة حشمت قاسم. ـ الرياض: مكتبة الملك فهد الوطنية، ١٤١٨هـ/ ١٩٩٧م.

٩- محمد فتحى عبد الهادى: (١٩٩٥) المصادر المرجعية للمعلومات فى العلوم الاجتماعية. ـ القاهرة: دار غريب للطباعة والنشر والتوزيع، ٣٩١ ص.

١٠- ناريمان إسماعيل متولى: (١٩٩٨) التكشيف والاستخلاص الآلى: أضواء من الإنتاج الفكرى الأجنبى خلال أربعين عاماً. **مجلة المكتبات والمعلومات العربية**؛ س ١٨، ع٣، ص ٦٢-٨٩ (الجزء الأول)، س ١٨، ع٤، ص ١٩-٤٣ (الجزء الثانى).

١١- ناريمان إسماعيل متولى: (يوليو ١٩٩٨) المستخلصات والاستخلاص، دراسة تحليلية تقييمية ــ **مجلة الاتجاهات الحديثة فى المكتبات والمعلومات**، س١٠، ع١٠ ص ٧٥-١١٤.

١٢- ناريمان إسماعيل متولى: (١٩٩٧). تكنولوجيا النص التكوينى (الهيبرتكست) وتنمية الابتكار لدى الطلاب والباحثين. ــ مجلة المكتبات والمعلومات العربية. ــ س ١٧، ع١ (يناير ١٩٩٧) ص ص ٥-٣٥.

١٣- يسرية زايد: (١٩٩٤). المستخلصات وأساليب الاستخلاص. ــ الاتجاهات الحديثة فى المكتبات والمعلومات. ــ ع٢، ص ص ٨٨ ــ ١٠٧.

المراجع الأجنبية

1- Abu- Salem, H. (1992) A micro computer based Arabic bibliographic information retrieval system with relational thesauri. ***Unpublished Ph. D. Dissertation. Chicago: Department*** of computer Science, Illinois Institute of Technology, Chicago.

2- Airchison, J (1970) Thesaurofacet: A Multipurpose Retrieval Language Tool. ***Journal of Documentation*** V. 26, P. 187- 203.

3- Aitchison, Jean (1997) ***Thesaurus Construction and Use: a Practical Manual*** / Jean Aitchison; Alan Gilchrist; Daird Barwdcn. 3rd ed. London: Aslib.

4- Aladesulu, O.S. (1986) Improvement of automatic semantically equivalent syntactically different phases. ***Dissertation Abstracts international,*** 46 (9), 3120- B.

5- Albright, J. B. (1979) Some limits to subject Retrieval from a large published Index. Doctoral Thesis. Urbana: Univ. of Illinois, Graduate School of library Science.

6- Al-Fedaghi, S, and Al- Anzi, F.(1989) A new algorithm to generate Arabic root- pattern forms. Proceedings of the 11th National Computer Conference and Exhibition, March (p. 391- 400). Dhahran, Saudi Arabia: King Fahd University of pctroleum and Minerals.

7- Al-Kharashi, Ibrahim and Evans, Martha (1994). Comparing Words, Stems, and Roots as Index Terms in an Arabic Information Retrieval System. ***JASIS,*** V. 45 (8): 548- 560.

8- Allan, J. (1995) Automatic Hypertext Construction. Ph. D. Dissertation (Technical Report TR 95- 1484., Computer Science Department, Cornell University, Ithaca, N. Y.

9- Alterman, R. (1992) Text summarization. In: Encyclopedia of Artificial intelligence (PP 1579- 1587). New York: Wiley.

10- American National Standards Institute Inc. (1979) American National Standard for Writing Abstracts. ANSI, Z 39.14 - 1979, New York: American National Standards Institute, 1979.

11- Anderson, J. D. (1986) Indexing System; Extensions of the Mind's Organizing power. ***Information and Behavior*** v.1, p. 287- 323.

12- Attar, R.; Fraenkel, A. S. (1981) Experiments in local metrical feedback in full- text retrieval systems. ***Information Processing and Management***, V. 17, P. 45- 126.

13- Austin, d.; Dykstra, M. (1984). PRECIS: A manual of concept analysis and subject Indexing. 2nd London: The British.

14- Bakewell, K. G. B. (1987). Reference Books for Indexers. ***The Indexer***, 15 (3): 131- 140.

15- Bates, M. J. (1986) Subject access in on- line catalogs: a design model. ***JASIS***, 367- 376.

16- Baxendale, P. B. (1958) Machine-made index for technical literature: an experiment. ***IBM Journal of research and Development*** V. 2, P. 354- 361.

17- Bearman, T. c. (1980) Indexing and Abstracting in: ALA World Encyclopedia of library and Information Services. Chicago: ALA.

18- Belkin, N. J. and W. B. Croft (1992) Information ***Filtering*** and Information Retrieval: Two Sides of The Coin. ***ACM Commun***. 35 (12), 29-38.

19- Bhatta charyya, G. (1981) Elements of Popsi. In: Indexing Systems: Concepts, Models and techniques. ed. by T. N. Rajan pp 73- 102. Calcutta, Indian Association of Special libraries and Information.

20- Borko, H, and Bernick, M. (1963) Automatic document classification. ***Journal of the Association of computing Machinery***, V. 10, p. 151- 162.

21- Borko, H., Bernier, C. (1975). ***Abstracting Concepts and Methods***, New York: Academic press.

22- Borko, H (1977). Toward A theory of Indexing. ***Information Processing and Management***, 13, 355- 365.

23- Brandow, R.; Mitze, K. and Rau, L. F. (1995) Automatic Condensation of Electronic Publications by Sentence selection. ***Information Processing and Management***, 31 (5), p. 675- 685.

24- Brenner, E. H. (1989). "Should Abstractors Index"? ***Newsletter of the American Society of Indexers***, No. 91, (March/ April).

25- Brookes, B. C. (1986) Jason Farradane and Relational Indexing.

Journal of Information Science, 12 (1/2), 15- 18.

26- Brown, K. et al (1990) A Statistical approach to machine translation. ***Computation Linguistics***, V. 16 (2), P. 77- 85.

27- Byrne, J. R. (1975) Relative effectiveness of titles, abstracts and subject headings for machine retrieval from the COMPENDEX Services. ***Journal of the American Society for Information Science***, V. 26, 223- 229.

28- Cardie, e. (1997) Empirical Methods in information ***Extraction. Al-mag*** 18(4), 65-79.

29- Chan, L. M; Richmond, P. A. Svenouis, E. (1985). ***Theory of subject Analysis: A Source book. Littleton,*** Colorado: Libraries Unlimited.

30- Chen et al, (1993) Generating, integrating and activating Thesauri for concept- based document retrieval. ***IEEE*** Expert, Special series on Artificial Intelligence in Text based information Systems, 8.25- 34.

31- Chen, H; Yim, T.; Fye, D. Schatz, (1995) Automatic Thesaurus Generation for an electronic Community System. ***Journal of the American Society for Information Science***, V. 46(3), P. 175- 193.

32- Church, K. (1988) A Stochastic parts Program and noun phrase parson for unrestricted text. In: **proceeding of the 2nd Conference on Applied Natural Language Processing.** PP. 136- 143. Austin., Texas.

33- Cleveland, D. B. and Cleveland, A. D. (1990) ***Introduction to Indexing and Abstracting- 2nd ed***, Englewood; Colorado: Libraries Unlimited, Inc.

34. Cleverdon, C. W. et al (1968) ***Factors Determining the Performance of Index Language*** (Cranfield, College of Aeronautics, 3 vols.

35- Cleverdon, C. W. (1977). A Comparative Evaluation of Searching by Controlled Language and Natural Language in an Experimental NASA Database. Frascati, European Space Agency, space Documentation Service.

36- Cochran, P. A. (1986) Improving LCSH for Use in online Catalogs: Exercises for Self Help with a Selection of Background Readings. Littleton, Co: libraries unlimited, 348p.

37- Coco, A. (1984) Full- text versus full- text plus editorial additions. ***Legal Reference Services Quarterly***; 4 (2), 27- 37.

38- Cohen, J. D. (1995) Highlights: Language and Domain independent Automatic Indexing terms for Abstracting **Journal of the *American Society for Information Science,*** 46 (3), 163- 174.

39- Collison, Rabort (1971). ***Albstracts and Albstracting Services.*** Santa Barbara, CA: ABC Clio.

40- Cowie, J. and W. Lehet (1996) Information Extraction ***ACM commun***. 39 (1), 80-91.

41- Craven T. C. (1986) String Indexing. Orlando, Fla: Academic Press.

42- Cremmins, E. T. (1996) The Art of Abstracting. 2nd ed. Arlington, VA: Information Resourses Press.

43- Croft, W. B. (1987) Approaches to Intelligent Information Retrieval. ***Information Processing and Management,*** 23 (4) 249- 254.

44- Crouch, e. J (1990) An approach to the automatic construction of global Thesauri. ***Information processing and management,*** V. 26 (5), P. 629- 640.

45- Diodato, V. P. (1981) Author Indexing in Mathematics. Doctoral thesis. Urbana: Univ. of Illinois, Graduate School of library and Information Science.

46- Dubois, C. P. R. (1987) Free text vs controlled vocabulary; a reassessment. ***Online Review,*** v. 11, 243- 253.

47- Dykstra, M. (1987) Subject Indexing and Retrieval: what more can technology Do? ***Canadian Library Journal,*** 44 (3): 187- 189.

48- Edmundson. H. P. Oswald V. Wyllys, R. (1959) ***Automatic Indexing and Abstracting of the contents of documents.*** Los Angels, CA.

49- Edmundson, H. P. (April 1964) Problems in Automatic Abstracting. ***Communications of the ACM,*** 7 (4), 259- 63.

50- El- Gihani, A., Sharp, B. (1995) Text understanding for automatic abstracting. Proceedings of the thirteenth INFORSID Congress, P. 189- 201, INFORSID, Toulouse.

51- Endres- Niggemeyer, B.; Maier, E. and Sigel, A. (1995). How to implement a Naturalistic Model of Abstracting. ***Information***

processing and Management, 31 (5), P. 631- 674.

52- *ERIC Processing Manual* (1982). Washington, D. C., U. S. Department of Education, Educational Resources Information Center.

53- Fayyad, V., Piatetsky-Shapiro, G., Smyth, P. (1996) From data mining to Knowledge discovery in data bases. A: Magazine, V. 17 (3), 37-54.

54- Fetters, L. K. (1987) A Guide to Indexing Software. 2nd ed. Washington, D. C. American Society of Indexers.

55- Fidel, R. (1986) Writing abstracts for free- text searching. *Journal of Documentation*, 42 (1), 11- 21.

56- Foskett D. J. (1963) Classification and Indexing in the Social Sciences. London: Butter-worth.

57- Foskett, D. J. (1971) Problems of Indexing and Classification in the social Sciences. *International Social Science Journal*, 23, no. 2, 244- 55.

58- Fugmann, R. (1985) The five- axiom theory of indexing and information supply. *JASIS*, 36, P. 116- 129.

59- Garfield, E. (1979). Citation Indexing. Its theory and application in science, Technology and Humanities. New York: Wiley.

60- Garfield, E. (1980) Is Information retrieval in the Arts and Humanities inherently different from that in science? The effect that ISI's citation from that index for the Arts and humanities is expected to have on future scholarship. *Library Quarterly*, 50, 40- 47.

61- Gazdar, G. and Mellish, C. (1989) Natural language Processing in LISP: Introduction to Computational Linguistics. Addison-Wesley.

62- Gillbert, N; Woolgar, S. (1974). The Quantitative study of Science: an examination of the literature. *Science studies*, 4, No. 3, 279- 294.

63- Gopinath, M. A. (1987) Symbiosis between Classification and Thesaurus. *Library Science with a slant to Documentation* 24 (4): 211- 225.

64- Grande, S. (1987) The Ideograph as Model for database design, Indexing and Retrieval of Geological Information. *Canadian*

Journal of Information Science 12 (1): 10- 19.

65- Grishman, R. (1986) Computational Linguistics: An Introduction. Cambridge University Press.

66- Hahn, V. and Schattings, K. (1997) Deep Knowledge discovery from Natural language texts. ***Proceedings of The Third International Conference on Knowledge Discovery and data mining.*** New port Beach, CA. 1. 175-178.

67- Hamphrey, S. M. and Miller, N, E. (march 1987). Knowledge based Indexing of the Medical literature: ***The Indexing*** Aid Project. ***JASIS***, 38, 184- 96.

68- Harbourt, A. M.; Knecht, L. S.; Humphreys, B. L. (1995). Structured abstracts in Medline, 1989- 1991. ***Bull. Med. Libr. Assoc.***, V. 83, No. 2, P. 190- 195.

69- Hartley J., Sydes, M(1996) Obtaining information accurately and quickly: Are structured abstracts more efficient? ***Journal of Information Science***, 22 (5), 349-356.

70- Harvey, L. (1982). The use and abuse of Kuhnian Paradigms in the sociology of knowledge. ***Science***, 16, No 1, 85- 101.

71- Hersey, D. F. et al (1971) Free text word retrieval and scientist indexing: performance profiles and costs. ***Journal of Documentation***, V. 27, 167- 183.

72- Hmeidi, Ismaill; Kanoan, G. and Evans, Martha (1997). Design and Implementation of Automatic Indexing for Information Retrieval with Arabic Documents. **JASIS**, V. 48 (10): 867- 881.

73- Hodges, Julia (2000) Automated System for the generation of Document Indexes. In: *ELIS* , V. 66 Supply. 29, P.1-19.

74- Horty, J. F. (1982) Legal Research using electronic techniques. In: Literature of law- Techniques of Access (Proceedings of the 5th American Association of law Libraries Institute for law librarians, M 56- 68) South Hackensack, N. J., F. B. Rothman and Co.

75- Hovy, E. (1991). Approaches to the Planning of coherent text. In: c. Paris, w. Swartout and W. C. Mann (eds) Natural language generation in artificial intelligence and computational Linguistics. Amsterdam kluwer Academic.

76- Hovy, E. (1993). Automated discourse generation using a discourse structure relations. ***Artificial Intelligence*** 63, 341- 385.

77- Hurt, C. D. (1998) Information Sources in Science and Technology. Englewood, libraries unlimited.

78- Hutchins, W. J. (1978) The Concept of "aboutness" in subject indexing ***Aslib Proceedings*** V . 30, p. 172-181.

79- Hutchins, W. T. (1987). **Summarization: Some problems and Methods.** In: K. P. Jones (ed.) Informatics 9: Meaning: The frantier of informatics (P. 151- 173). London: Aslib.

80- International organization for standardization (1982) Information transfer: 1S0 Standards Handbook I (2nd ed). Geneva, Switzerland.

81- Johansen, T. (1987) Elements of the Non- Linguistic approach to subject relationships. ***International classification***, 14 (1): 11- 18.

82- Johnson, F.(1995). Automatic abstracting research. ***Library Review***, 44 (8), p. 28- 36.

83- Jonak, Z. (1984) Automatic indexing of full text. ***Information Processing and Management***, 20 (5/ 6), 619- 27.

84- Keen, E. M. and Digger, J. A. (1972). Report of an Information Science Index Language Test. Aberystwyth, College of Librarianship, Wales, 2v.

85- Kent, A., Taulbee, O. E.: Belzer, J. & Goldstein, G. (eds) (1957) Electronic handling of Information: Testing and Evaluation (p.p 173-185). Washington D. C., Thompson Book Co.

86- King, R. A. (1976). A Comparison of the readability of abstracts with their source documents. ***Journal of the Amer. Soc. For Information Science***, V. 27, p. 118- 121.

87- Knapp, S. D. (1983) BRS/ TERM. Database for researcher online, 83 Conference Proceedings, p. 162- 166. Weston, Connecticut, Online Inc.

88- Koltay, T. (1995). The visible challenge: A hypertext tutorial on abstracting for library Science Students. ***Journal of Education for Library and information Science,***. V.36, No.2, P. 170- 177.

89- Krovetz, R. and Croft, W. B. (1989) Word sense disambiguation using machine readable dictionaries. In: Proceedings of the 12 the International S/ G/ R Conference on Research and

Development. In information Retrieval/ edited N. J. Belkin and W.B. Corft, p. 127- 137. Boston.

90- Kuhlen, R. (1984). Some similarities and differences between intellectual and machine text understanding for the purpose of abstracting. In: Representation and Exchange of knowledge as a basis of Information Processes ed by H. J. Dietschmann, p. 87- 109. Amsterdam. North Holland.

91- Kuhn, T. S.(1970). The Structure of scientific revolutions. Rev. ed. Chicago: University of Chicago Press.

92- Lancaster, F. w. (1968). Vocabulary Control for Information Retrieval. 2nd ed. Arlington, Vo., Information resources press.

93- Lancaster, F. W. Goldhor, H. (1981). The Impact of on- line Services on subscriptions to printed publications. ***On-line Review***, 5, P. 301- 311.

94- Lancaster, F. W.; Elliker, C.; Connell, T. H. (1989). Subject Analysis. ***Annual Review of Information Science and Technology*** (ARIST), Vol. 24, pp.35- 84.

95- Lancaster, F. (1991). Indexing and Abstracting in Theory and Practice. London: The Library Association.

96- Lancaster, F. W. (1998) Indexing and Abstracting in theory and practice, 2nd ed. Unbana – Champaign, Univ. Illinois.

97- Lancaster, F. W. & Smith, L. C. (1999) Intelligent technologies in library and information service applications: A realistic appraisal. Medford. N. J. Information today.

98- Langridge, D. W. (1976). Classification and Indexing in the humanities. London: Butter-worth& Co.

99- Larson, R.R. (1991). The Decline of subject searching: Long term trends and patterns of index use in on online Catalog. ***JASIS***, V. 42, P. 197- 212.

100- Learn, L. L. (1993). The Implications of recent telecommunication developments for abstracting and indexing services. In: Three views of the Internet, P. 5-20, National Federation of Abstracting and Information Services, Philadelphia, PA.

101- Leanard, I. E. (1975). Inter- Indexer Consistency and retrieval effectiveness: Measurement of Relationship. Doctoral thesis.

Urbana. Univ. of Illinois, Graduate school of library Science.

102- Leavy, M. D. (1983). Obliteration in the Natural and Social Sciences. Citation Data in Search of a theory. ***Int. forum inf. and Doc.***, V. 8, No 4, 27- 31.

103- Leung, C. H. Kan, W. K. (1997). A statistical learning approach to automatic indexing of controlled index terms. ***JASIS***, 48 (1), 55- 66.

104- Line, M. B. (1974). Concluding Considerations. "In: ***Access to the Literature of the Social Sciences and Humanities*** Flushing, N. Y. Queens College Press.

105- Loukopoulos, L. (1966). Indexing Problems and some of their Solutions. ***American Documentation.*** p. 17- 25.

106- Luhn, H. P. (oct. 1957). A statistical Approach to Mechanized Encoding and searching of literary Information. ***IBM Journal of Research and Development***, 1 (4), 309- 317.

107- Luhn, H. P. (April 1958). The Automatic Creation of literature Abstracts. ***IBM Journal of Research and Development***, 2, 159- 65.

108- Lunin, L. (1967). The Development of a machine. Searchable index- abstract and its application to biomedical literature. In: ***Three Drexel information Science Research studies***, ed. By B. Flood, PP 47- 134. Philadelphia, Drexel Press.

109- Markey, K. et al (1980) An Analysis of controlled Vocabulary and free- text search statements in online searches. ***Online Review***, 4, 225- 236.

110- Markey, k. (1987) Searching and Browsing the D. D. C. in an online catalog. ***Cataloging and classification Quarterly***, 7 (3): 37- 68.

111- Maron, M. E. (1988). Probabilistic design principles for conventional and full- text retrieval systems. ***Information processing and Management***, 24, 249- 255.

112- Martin, W. A. (1980) Toward an integral multi- file on- line bibliographic database. ***Journal of Information Science***, V.2, 241- 253.

113- Mathis, B. A. (1972). ***Techniques for the Evaluation and Improvement of computer- produced Abstracts.*** Columbus,

Ohio state university, computer and Information Science research center.

114- Maybury, M. T. (sep 1995). Generating summaries form event data. ***Information processing& Management*** 31 (5), P. 735- 51.

115- Mc Coin et al (1987). Comparing retrieval performance in online database. ***Information processing and Management***, 23, p. 539- 553.

116- Mc Keown, K.; Robin, J.; Kukich, K. (1995) Generating concise natural language summaries. Information Processing and Management, Vol. 31, No. 5, 703-733.

117- Milstead, J. (1994) Needs for Research in Indexing. ***JASIS***, v. 45(8), p. 577-582.

118- Moffat, A. (aug. 1995). In site generation of compressed inverted files. ***JASIS***, 46 (7), p. 537- 50.

119- Molina, M. P.(1995). Documentary Abstracting: Toward a Methodological Model. ***Journal of the American Society for Information Science***, Vol. 46, No. (3), P. 225- 234.

120- Morris, A. H.; Kaspar, G. and Adams, D. (1992). The effects and limitations of automated text condensing on reading comprehension performance. ***Information systems research***, 3 (1), 17- 35.

121- Naisbitt, John (1984) Mega trends. New York, Warner Books: Naisbitt

122- National Information Standards Organization (1997) Guidelines for Abstracts. Bethesda, MD: NISO.

123- O'Leary, M. (1990) Producing a data bass: Many Choices for Data entry. ***Data Bass***. V.38, p. 343-435.

124- Paice, C. (at. 1986). Expert systems for Information Retrieval. ***Aslib proceeding***, 38, 343-435.

125- Paice, C. (1990). Constructing literature abstracts by computer: technique and prospects. ***Information processing and management***, 26 (1), 171- 186.

126- Perez, E. (1982). Text enhancement: controlled vocabulary vs free text special. libraries, V. 73, 183- 192.

127- Perry, J. W. and Kent, A. (1951) ***Documentation and Information***

Retrieval. Cleveland, Press of W R U.

128- Perry, J. W. & Kent, A. (1958) ***Tools for machine Literature searching.*** New York: Inter-science Publishers Inc.

129- Pinto, M and Lancaster, F. W. (Summer 1999) Abstracts and Abstracting in Knowledge Discovery. ***Library*** **Trends,** V. 48 (1), 234-248.

130- Price, D. J. de s. (1970). Citation Measures of hard science, soft science, technology and non science. In. C. E. Nelson and D. K. Pollok (eds) ***Communication among scientists and engineers.*** Lexington, Mass: Heath- Lexington, 3- 22.

131- Raghaven, V. V., Deogun, J. S. Sever, H. (eds) (1998) Introduction (In Special Topic Issues: Knowledge Discovery and data Mining) **JASIS,** V. 49 (5), 397-402.

132- Rasheed, M. A. (1989). Comparative Index terms. ***International Library Review,*** 21, p. 289- 300.

133- Rolling, L. (1981). Indexing consistency, quality and efficiency. ***Information Processing and Management.*** 17, P. 69- 76.

134- Rothkegal, A. (Sep. 1995). Abstracting from the perspective of text production. ***Information Processing and management,*** 31 (5), P. 777- 84.

135- Rowley, J. E. (1988). ***Abstracting and Indexing.*** 2nd ed. London: Clive Bingley.

136- Salton, G. (June 1971). Automatic Indexing using Bibliographic Citation. ***Journal of Documentation,*** 27, 98- 110.

137- Salton, G. (1975). ***Dynamic Information and library Processing.*** New Jersey: Prentice Hall, Inc. (Automatic Abstracting, p. 106- 109, automatic Indexing, p. 76- 106.)

138- Salton, G: Yang, C. S.; Yu, C. T. (1975) A theory of term importance in automatic text analysis. JASIS, 26 (1), p. 33- 44.

139- Salton, G. and Mcgill, M. (1983). ***Introduction to Modern Information Retrieval.*** New York: Mc Graw- Hill.

140- Salton, G. (1986). Another look at automatic text- retrieval systems. ***Communications of the ACM,.*** 29, 648- 656.

141- Salton, G. and Buckley, C. (1988). Term- Weighting Approaches in Automatic Text Retrieval. ***Information Processing and***

Management, 24 (5), 513- 523.

142- Salton, G. (1989). ***Automatic text Processing. Reading***, MA: Addison- Wesley.

143- Salton, G.; Singhal, A.; Buckley, C. and Mitra, M. (1996). Automatic text decomposition using text segments and text themes. Hypertext 96 The seventh ACM Conference on ***Hypertext*** (p. 53- 65). New York: Association for Computing Machinery.

144- Salton, G.: Singhal, A.: Mitra, M. and Buckley, C. (1997). Automatic Text Structuring and Summarization. ***Information Processing and Management***, 33 (2), P. 193- 207.

145- Schnelling, H. M. (1986). Pattern Indexing: Towards universal structures and Transparency of Indexing: Literary Scholarship as a case in point. ***Cataloging and classification Quarterly***, 7 (1), 35- 44.

146- Smeaton, Alan F. (1992) Progress in the Application of Natural language Processing to Information Retrieval Tasks. ***The computer Journal***, V. 35, (3), 268- 278.

147- Sparck, J. K. (1981). ***Retrieval system* tests** 1958- 1978. Sparck Jonesk. ed. Information Retrieval experiment London: Butteworks.

148- Svenonus, E. (1981). Directions for research in indexing, Classification and cataloging. ***Library Resources and Technical services***, 25. 88- 103.

149- Taylor, Robert (1984). Value- added Processes in Document-based systems: Abstracting and Indexing Services. ***Information Services and use***, V. 4. p. 127- 46.

150- Tenopir, C. (1984). Retrieval Performance in a Full text. ***Journal Article Database***. Doctoral Thesis. University of Illinois, Graduate school of library and Information Science. Also in ***on- line Review***, V. 9, 1985, 149- 164.

151- Tenopir, Coral; Jacso, Peter. (1993). Quality of Abstracts. ***On line.*** 17 (3). P. 45- 55.

152- Tian-long Wan; Martha Evans; Yeun. Wen Wan; Yuen-Yuan pao (1997) Experiments with Automatic Indexing and a Relation Thesaurus in a Chinese Information Retrieval System. ***JASIS***, vol. 48, No. 12, pp. 1086-1096.

153- Tibbo, H. R. (1992). Abstracting across the Disciplines: a content analysis of abstracts from the natural science, the social science and the humanities with implication for abstracting standards and on- line information retrieval. ***Library and Information Science Research***, Vol. 14, No. (1), P. 31- 56.

154- Tinker, J. F.(1968). Imprecision in indexing. ***American Documentation***, 19, P. 322- 330.

155- Tomaiuala, N. G. and Jacqueline Trolley (1992). Science citation. Index/ Science Citation Index with Abstracts on CD. ROM. ***CD. ROM librarian.***- Vol. 7, No. (9), P. 34- 41.

156- Tomaiuala, N. Q. (1993). Social works Abstracts plus on Compact disc. ***CD- ROM World.*** Vol. 8, No (7), P. 79- 82.

157- Tothkegal, A. (Sep. 1995) Abstracting from the perspective of text production. Information Processing and Management, Vol. 31, No. 5, pp. 777-84.

158- Vinsonhaler, J. F. (1996). Some behavioral indices of the viability of document abstracts, ***Information Storage and Retrieval***, Vol. 3, P. 1-11.

159- Wanger, J. et al (1980) Evaluation of the on- line process, Santa Monica, Cuadra Associates, PB81- 132565.

160- Waters, S. (Sept. 1986). Answer man, the Expert Information Specialist. An Expert System for retrieval of information from library reference Books. ***Information technology and libraries***, 7, 205- 11.

161- Weil, B. H. et al (1963). Technical Abstracting fundamentals. ***Journal of chemical Documentation.*** Vol. 3, P. 86- 89, 125- 136.

162- Weil, B. H. (1970). Standards for Writing abstracts. ***Journal of the American Society for Information Science***, P. 351- 357.

163- Weinberg, B. H. (1988). Why Indexing Fails the Researcher ***The Indexer*** V. 16, No. 1, 3-6.

164- Wheatley, A. Armstrong, C. J. (1997). Metadata, recall and Abstracts Can Abstracts ever be reliable indicators of document Value? ***Aslib Proceedings***, V. 49 (8), 206-213.

165- White, H. D.; Griffith, B. C. (1987). Quality of Indexing in on-

line Data Base. ***Information Processing and Management,*** 23 (3): 211- 224.

166- Wiberely, S. E. (1983). Subject Access in the Humanities and the Precision of the Humanist's Vocabulary. ***Library Quarterly,*** V. 53, No 4, 420- 433.

167- Wiley, D. L. (1994). Can the traditional abstracting and indexing Services Survive? ***Data base.*** Vol. 17, No. (6), P. 18- 24.

168- Winograd, P. N. (1984). Strategic difficulties in Summarizing texts. ***Reading Research Quarterly,*** 19, 404- 425.

169- Witty, F. J. (1973). The Beginnings of Indexing and Abstracting: Some Notes toward ***History of Indexing and Abstracting*** in Antiquity and the Middle ages. ***The Indexer,*** V. 8, p. 194- 198.

170- Woods, S. E. (1994). Current Contents on disk with abstracts. Medical reference Services Quarterly, Vol. 13, No. (4), P. 45- 54.

171- Yeates, R. (1988). Prestel Indexing from the user's Point of view. ***The Indexer,*** 16 (1): 7- 10.

172- Yu, K. I. Et al (1987). Pipelined for speed: the fast data finder system. Quest (TRW Defense and space Systems Group), 9 (2), 5- 19.

173- Zarri, G. p. (1983). An outline of the representation and use of temporal data in the RESEDA System. ***Information Technology Research and Development*** V. 2, P. 89- 108.

174- Zholkona, A. I. (1975). Applying facet analysis methods in abstracting. ***Scientific and Technical Information Processing,*** Vol. 2,P. 70- 74.



التكشيف والاستخلاص

هذا أحدث كتاب فى التكشيف والاستخلاص ظهر فى مطلع القرن الحادى والعشرين ويتناول أساسيات المجال ، فضلا عن التطورات والبحوث الجارية من الناحيتين النظرية والتطبيقية ، وإذا كان المؤلفون قد اهتموا بدراسة تاريخ التكشيف والاستخلاص وتعاريفه وأهميته فى استرجاع المعلومات ، فقد ألقوا الضوء على الاتجاهات المستقبلية التى يتوقعها المختصون فى المجال.

وفى فصول متميزة تناول المؤلفون مستويات التكشيف وأنواعه واللغة الطبيعية والمحكومة فى استرجاع المعلومات ، والمكانز وتحليل الإنتاج الفكرى العربى ، وتقييم ومقارنة التكشيف والاستخلاص الآلى والإنسانى ، مع إبراز تفاعلهما - فى الإنتاج الفكرى حتى عام ٢٠٠٠ م - مع الحاسبات الآلية واللغويات والذكاء الاصطناعى والنظم الخبيرة وغيرها ...

أى أن الكتاب يتوجه إلى طلاب المراحل الجامـ
فى أقسام المكتبات والمعلومات بالجامعات العربية ،
للباحثين وأعضاء هيئة التدريس العاملين فى هذا الحق

أحمد غريب

